DOCUMENT RESUME

ED 373 547 FL 022 249

AUTHOR	Smith, James Brian
TITLE	A Study of Item Bias in the Maine Educational Assessment Test.
PUB DATE	May 93
NOTE	167p.; Doctoral dissertation, Boston College.
PUB TYPE	Dissertations/Theses - Doctoral Dissertations (041)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	*Bilingual Students; Comparative Analysis; Elementary Secondary Education; English; French; Grade 8; *Mathematics Tests; Monolingualism; *Reading Tests; *Standardized Tests; State Programs; Statistical Analysis; *Test Bias; *Test Items; Test Validity
IDENTIFIERS	*Maine Educational Assessment

ABSTRACT

A study used four statistical item bias analysis strategies to determine the French cross-cultural validity of the Maine Educational Assessment, a standardized test administered in si.. content areas to students in grades 4, 8, and 11. Analysis was performed on eighth grade pupil performance in test year 1988-89, in the areas of the 100 common reading and mathematics items that all pupils take. The four statistical procedures used were: Scheuneman's modified chi-square; Rudner and Convey's TID-45 degree item difficulty p-value; the Rasch one-parameter latent trait model; and the Mantel-Haenszel procedure. Item response comparisons were made with two of Maine's pupil populations: 336 French bilingual/English-fluent speakers and 336 monolingual English-speakers. Findings show that 8 of 50 mathematics items and 9 of 50 reading items indicated differential validity. However, they also indicated differential functioning in not favoring significantly either language group. It is concluded that lower mean scores on the test among bilingual/English-fluent students are not due to item bias, as suggested in an earlier study. It is suggested that misidentification of bilingual/English-fluent students is a primary cause of the discrepancy in scores. (MSE)

A STUDY OF ITEM BIAS IN THE MAINE EDUCATIONAL ASSESSMENT TEST

by

JAMES BRIAN SMITH

Boston College

The Graduate School of Arts and Sciences

Department of Education

A STUDY OF ITEM BIAS IN THE MAINE EDUCATIONAL ASSESSMENT TEST

a dissertation

by

JAMES BRIAN SMITH

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Education

May 1993

BOSTON COLLEGE

GRADUATE SCHOOL OF ARTS & SCIENCES

The thesis of James Brian Smith

entitled A Study of Item Bias in the Maine Educational Assessment Test

submitted to the Department of Education

in partial fulfillment of the requirements for the degree of Doctor of Education

in the Graduate School of Arts and Sciences has been read and approved by the Committee:

Vincent C Rucci

Barry Berube

Peter W. Airasian

Date January 29 1993

ABSTRACT
A Study Of Item Bias In The
Maine Educational Assessment Test
J(ames) Brian Smith

This study used four statistical item bias analysis strategies, and a panel of French bilingual/English experts to determine French cross-cultural validity of the Maine Educational Assessment test (MEA), a program which grew out of the Maine Educational Reform Act of 1984. It is administered to Maine Students in grades 4, 8, and 11, in the subject areas of reading, mathematics, writing, science, social studies, and humanities.

An analysis was done of eighth grade pupil performance in test year 1988-89, in the areas of the 100 common reading and mathematics items that all pupils take. The statistical strategies used were:

Scheuneman's modified chi-square procedure (SSX^2);

Rudner and Convey's TID-45^0 item difficulty p-value;

The Rasch one-parameter latent trait model;

The Mantel-Haenszel procedure.

Item response comparisons were made of two of Maine's pupil populations: 336 French bilingual/English fluent speakers from the communities of St. Agatha, Madawaska, Lewiston, Ft. Kent, Van Buren, and Caribou, and 336 monolingual English speaking pupils randomly selected and paired to insure identifiably equal abilities on the basis of equal overall test scores.

Findings indicated that eight out of fifty mathematics items (16%), and nine out of fifty reading items (18%) from the common questions asked of all eighth grade pupils in the two subject areas analyzed indicated differential validity. However, the identified items also indicated differential functioning in that they did not favor either language group in a material way. In fact this differential functioning finding suggests that the French bilingual/English fluent minority children have virtually equal chances of achieving identical scores on the common mathematics and reading questions as did the monolingual English pupils.

Lower mean scoring by bilingual/English fluent pupils, as indicated in the "1988-89 Maine Educational Assessment Grade 8 Mean Comparisons," is evidently not caused by item bias in the MEA as far as the French bilingual/English fluent pupils are concerned. Either the French pupils are not affected as are other cross-cultural groups, or other reasons are at play as to the origin of the lower means of Maine's French bilingual pupils. As indicated in a survey done in preparation for this study, mis-identification of bilingual/English fluent pupils is surely a primary cause.

ACKNOWLEDGEMENTS

I wish to thank the members of my dissertation committee: Dr. Vincent C. Nuccio for his untiring support throughout my program and for chairing my committee; Dr. Peter W. Airasian for his yeoman's service in leading me down the tortured path from polemic to dissertation; Dr. Barney A. Bérubé for his backing and amity during the nearly five years spent on this undertaking.

I want to thank Dr. Stuart Kahl, Vice President of Advanced Systems, for providing the generous and abundant services of Dr. Charles A. DePascale, Supervisor of Data Analysis & Reporting for Advanced Systems; and Charles DePascale for his assistance, patience, and skilled advice with the statistical analysis. All of the assistance from Advanced Systems was gratis in the interest of science and better assessment. I am deeply appreciative.

I am highly appreciative for the volunteer services of my Panel of Experts: Gilbert J. Albert, of Ft. Kent, ME., Karen Michaud, of Upper Frenchville, ME., Barney A. Bérubé, of Augusta, ME., Gilman Hébert, of Jamaica Plain, MA., Yvon A. Labbé, of Orono, ME., and Guy Roy, of St. David, ME.. Without these great people, this study in its present configuration would not have been possible.

I want to thank my family: my wife, Negar, my daughter Negin, and my three sons Vahid, Tristan, and James for their unstinting support in this seemingly interminable

undertaking. For nearly five years they were largely without husband and father. I will try to make it up!

I want to thank my parents Clifford and Ruth Smith and my aunt and uncle Frederick and Frances Schipper for their encouragement and financial support with this undertaking.

I want to thank my mother-in-law, Fatemeh Borhan, of Shiraz, Iran, for expecting me to do this doctorate.

Finally, I want to thank H. Johnson Nenty, Ph.D of the University of Calabar, Calabar, Nigeria, for graciously giving me permission to quote extensively from his work in the process of replicating his statistical methodology.

Dedication

With esteem and affection, I dedicate this dissertation to the Passamaquoddy Indians of the Indian Township and Pleasant Point Reservations, and to the Penobscot Indians of the Indian Island Reservation, all of Maine. Most especially, I want to thank the members of the three school committees who from the outset have supported me with released time and with financial assistance throughout this study. Also, specifically, I want to credit the late J. Hartley Nicholas, former Governor of the Pleasant Point Reservation, who planted the seed which got me started on this research.

It is my fervent hope that the work started here will provide the impetus for better serving the educational needs of the children of these three great Peoples.

TABLE OF CONTENTS

List of Tables

ILLUSTRATIONS

CHAPTER ONE

THE PROBLEM AND PURPOSE OF THE STUDY

PROBLEM TO BE INVESTIGATED

Given the consistently lower mean test performance among State of Maine bilingual/English fluent pupils, compared to monolingual English pupils, the results of the Maine Educational Assessment in reading and mathematics were analyzed to determine whether there was bias in item structure and format .

The Maine Educational Assessment (MEA) Program grew out of the Maine Educational Reform Act of 1984. That Act called for a comprehensive set of reforms directed toward school improvement.

The goals of the MEA assessment program as mandated by the legislature were--

- o provide information on the academic achievement and progress of Maine students;
- o establish a process for continuing evaluation of state educational goals and aid in the development of educational policies, standards and programs;
- o provide school officials with information to assess the quality, effectiveness and appropriateness of educational materials, methods and curriculum needs, including remediation and enrichment;
- o provide school staffs with information about individual students which may be used, with other

information, to meet individual educational needs of the student;

- o identify year-to-year trends in student achievement; and
- o provide parents with information about the achievement of their children on the assessment tests.

The Maine Educational Assessment combines aspects of standardized achievement testing and program assessment. A set of "common questions," administered to all students at a grade in the state, are intended to yield reliable individual student test and subtest scores in reading, writing, and mathematics. The common item set, however, can not provide the broad coverage of content areas and program evaluation capability afforded by other tests. Therefore, other questions, called "matrix-sampled" questions, are distributed over many booklets so that each of these questions will be answered by a sample of students. Matrix-sampled questions are used in reading, mathematics, science, social studies, and the humanities to obtain a general indication of how pupils in Maine are performing in these areas. Many different test forms are used at each grade level tested; each containing the common core items and a fraction of the matrix sampled questions. The matrix-sampling technique allows for broader assessment of many content areas at the school level using a minimum amount of

testing time. The common core items are first tested in the matrix-sampling procedure before they are screened for possible use as a common item. Once used in the common sets they are discarded.

The tests are tailored to educational objectives developed by various advisory committees of Maine teachers, administrators, and curriculum experts. All skills areas except writing are tested with multiple choice questions. Writing is tested directly through writing prompts which elicit samples of students' writing. In addition, students are administered up to twenty open-ended reading and mathematics questions, which are intended to measure higher order thinking skills.

Student, teacher, and principal questionnaires assess a variety of background, experiential, attitudinal, and instructional variables. The information obtained from these instruments plays an important role in reporting and interpreting the assessment results. [1]

PURPOSE OF THE STUDY

The purpose of this study is to compare the test item responses of Maine's eighth grade pupils who are French bilingual/English fluent and eighth grade English monolingual pupils to determine whether student performance in the test year 1988-1989, on the common core of 50 reading

[1] Guide To The Maine Educational Assessment 1988-89; Department of Education, Division of Educational Assessment, 1-2.

and 50 mathematics test items indicate characteristics of differential validity defined in this study as item bias.

Students are reported as bilingual/English fluent if there is evidence that a language other than English is used in the student's home environment and that the student is proficient in all English communicative skills areas of reading, writing, speaking and listening commensurate with his/her monolingual English peers. It was intended that the Department of Education-administered Lau home language surveys would be used to determine if students were bilingual or monolingual. A student is reported as monolingual English if there is no evidence that a language other than English is used in the student's home environment. The Department of Education-administered Lau home language surveys were used to determine if students were bilingual or monolingual. As will be noted in Chapter II, these intended procedures have not been consistently carried out.

BACKGROUND OF THE PROBLEM TO BE INVESTIGATED

Disparities between the performance of bilingual/ English fluent and monolingual English test takers on the 1989 MEA are apparent. (See Table 1) "[Pupils] reported as bilingual who are fluent in English performed, overall, below their monolingual English peers." [2] The present

---

[2] Barney Bérubé, "Data Collection Report On Language Minority Children," Dept of Ed. Federal Projects for Language Minorities, Augusta Maine, October 1991, 46.

research looks for indications of biased test items as they may relate to the performance of the French bilingual/ English fluent pupil population of Maine's six major Franco American communities on the Maine Educational Assessment. The purpose of the study is to try to determine if test item bias is a reason for the lower mean scores of this minority language population.

The MEA results reported in 1989, show that the bilingual/English fluent pupils of the state, French (59.7%), Asian (18%), German (3%), Spanish (8.6%), American Indian (2,7%), and Other (8%), have lower mean scores than their monolingual English peers. It is unclear what is triggering lower mean scores in the group of bilingual/English fluent, but it is evident from their scores that these language minority pupils not only perform less well than all students statewide who took the MEA in these test years, but also performed less well than their monolingual English counterparts. Table 1 following documents this phenomenon.

TABLE 1					
1989 MEA Grade Eight Mean Comparisons [3]					
Mono	275	250	300	275	275
Bi	225	225	275	200	225
	Reading	Writing	Math	Science	S.S.

Mono = Monolingual English Speakers
Bi = Bilingual/English Fluent Speakers

Such a phenomenon is alarming in light of the fact that there would seem to be no evidence that use of a non-English language in the home is in itself a cause of failure in an all-English school environment. [4] Moreover, where a second language is used in the home, students normally tend to outperform their monolingual peers in a variety of subject areas. [5] Jensen makes a related claim..."Mexican-American children from bilingual homes (i.e., Spanish and English spoken by the parents) generally perform better on standardized tests than do children from homes in which Spanish is spoken exclusively." [6]

Hakuta, in discussing language acquisition issues

---

[3] Ibid., 49.

[4] J. Cummins, Schooling and Language Minority Students: A Theoretical Framework The role of primary language development in promoting educational success for language minority students. (Los Angeles: Office of Bilingual/Bicultural Education, California State Department of Education, California State University, Evaluation, Dissemination, and Assessment Center, 1981).

[5] M. Dornbusch, Bilingual Children Associated With Better Grades, East/West, 30:12, March 20, 1988, p.23.

[6] Arthur R. Jensen, Bias in Mental Testing, (New York: The Free Press, 1980), 606.

(Japanese by English Speakers), also concludes that:

> Comparisons of bilingual and monolingual children, as well as comparisons of bilingual children of varying levels of development, indicate that bilingualism can lead to superior performance on a variety of intellectual skills.... [T]here is widespread agreement among researchers that these effects are real, and there is overwhelming rejection of earlier research suggesting negative intellectual consequences of bilingualism. [7]

This research will look only at MEA test data for 8th graders tested in 1988-1989, who have been designated French bilingual/English fluent from the Maine communities of Caribou, Ft. Kent, Lewiston, Madawaska, St. Agatha, and Van Buren. It will attempt to discover if item bias is a factor which might explain why this language minority group, i.e., those both self-selected and administrator-selected as bilingual/English fluent speakers, is performing as it is on this state-wide instrument. The researcher will examine the test in part for possible bias through item analysis, and attempt to identify characteristics of the test where this bias may be occurring. The underlying motivation for conducting this research is to enter the debate about bias and test use which, inexorably, should lead to enhanced understanding of the broader concerns of social justice and the appropriate use of tests for groups affected by testing. [8]

---

[7] Kenji Hakuta, "Second Language Acquisition, Bilingual Education, and Prospects For A Language-Rich Nation," [draft], AD [Photostat], 1990, Stanford University, California; location of original unknown.

[8] Robert L. Linn, ed. Educational Measurement Third Edition, (New York: Collier Macmillan, 1989), 202.

Furthermore, as Jensen points out:

> [i]f the results of testing are of importance to the individual, the bilingual child should be tested in <u>both</u> languages by an [examiner] who is fluent in the [pupil's] primary language and its particular localisms, and the test should be scored in terms of the total number of correct responses in <u>either</u> language, with proper corrections for guessing, if the answers are multiple choice. [When such double testing is carried out...] maximum score[s] attained in both languages are usually not more than 5 to 10 points higher than in either language alone, but occasionally the difference is considerably greater, thus making this precaution worthwhile if any important decision[s] concerning...individual child[ren are] to be based on the test results. [9]

Surely Jensen's conclusions here can be extended from affected individuals to affected groups of individuals such as the French, Asian, German, Spanish, and American Indian groups in Maine. Decisions, especially those about school accountability with respect to teacher effectiveness and curriculum, are being made on the basis of the MEA tests by all manner of people: Department of Education consultants, the press, parents, the general public, the ethnic communities affected, and ultimately the children dominated by such testing. Certainly, if such is the case, then all parties involved have the absolute right to expect and to be assured that decisions about the development of test instruments are at least neutral with respect to how they are dealt with by the different cultural and linguistic subgroups being tested; and furthermore, that decisions to

---

[9] Jensen, 607.

provide public (media) generalizations or policy decisions affecting school accreditation and funding or even in extreme cases, program placement and/or curriculum modification, require forthrightly acknowledging the unique differences of the subgroups being tested, and the addressing of possible test item bias as a factor impacting the scores of differing subgroups.

Dolson argues that:

> ...[a] decision by [education] officials to abandon efforts to develop [equitable assessment] programs and policies...in favor of sociopolitical and economic concerns should not be made without great caution...[10]

Carried to the level of litigation, parties who feel aggrieved by the unconsidered use of such tests as the MEA, have legal precedence to call on for redress: see DIANA et al. v. State BOARD OF EDUCATION (California).

> An important aspect of the Diana case...is the contention that placement of children in special classes on whatever basis of assessment is in effect a form of labeling, which carries a stigma....[T]his issue of stigma takes on legal importance when the methods of assessment used to determine educational classification are called into question. Contention that a standard test is invalid for a certain class of children is therefore a basis for claiming that such children, if educationally classified by such tests, have been deprived of their rights to a proper education. This has been legally interpreted as a violation of the 'due process' and 'equal protection' clauses of the Fourteenth Amendment of

---

[10] David P. Dolson, "Bilingualism and Scholastic Performance: The Literature Revisited," The Journal of the National Association for Bilingual Education, 10, no. 1 (Fall 1985): 18.

the United States. [11]

In Maine, it is clear that at the very least value judgements about the quality of school departments serving the Indian minority population are taking place because of media handling of MEA test results. Because the three Indian community schools consistently score on the bottom of the score bands of the MEA statewide, it has become commonly accepted that something is "wrong" with Indian students, or with Indian schools. The test instrument itself has not been called into serious question by those making such assumptions. However, as has been noted earlier, this phenomenon of non-parity between the monolingual English pupils of the state, and the children who are bilingual/English fluent causes great concern to those concerned with education in these minority communities or communities that have a large language minority population. With respect to Indian Education, the number of pupils tested is so small that statistical results of an item bias analysis study, undoubtedly, would be called into question. Therefore, this study will focus on the much larger French communities where numbers are significant enough to warrant conclusions which may grow out of a study of item bias analysis, and where it is comparatively easy to identify the language minority population. Comparisons may then be drawn which could be pertinent to other cross cultural language

---

[11] Jensen, 31.

groups in the state...(as of October 1989, there were forty-nine such language groups reported, whereas, in 1993, the number has climbed to nearly 100); although, it is recognized that generalizations from one ethnic group to another can be problematical. [12] This study is perhaps all the more timely, as politicians currently consider mandating national testing similar to Maine's MEA tests.

TESTING BIAS

Bias, in the context of this study, will mean the presence of some characteristic of an item that results in differential performance of two individuals/groups of equal ability but from different subgroups.

Hambleton and Rogers suggest that bias can involve a.) sex, culture, ethnic, class and religious factors; b.) content; c.) language; d.) item structure and format; and e.) test time limits. [13]

This research will look only at the content, language, and item structure, and format classifications of bias as they may impact items in the MEA. It will look for differences between groups, and for cultural and language (non-ability) explanations of differences. Cole and Moss in discussing bias, have said that:

---

[12] Barney Bérubé, "Data Collection Report On Language Minority Children," Dept. of Ed Federal Projects for Language Minorities, Augusta Maine, 1989, 6-8.

[13] Ronald K. Hambleton, and H. Jane Rogers, "Design of an Item Bias Review Form: Issues and Questions," New York State Education Department, Albany (1988).

> [a]n inference from a test score is considered sufficiently valid when a variety of types of evidence supports its plausibility and eliminate primary counter inferences. An inference is biased when a test score has meanings or implications for a relevant, definable subgroup of test takers that are different from the meanings or implications for the remainder of the test takers. Thus, bias is differential validity of a given interpretation of a test score for any definable, relevant subgroup of test takers." [14]

In summary, "[p]sychometric bias is a set of statistical attributes conjointly of a given test and two or more specified subpopulations."[15] Bias is an error in measurement (unreliability), and in prediction (invalidity) both of which are related to the individual's group membership, and which exists when individuals are in some way discriminated differently than the criterion measure of performance is designed to discriminaate. Of course the criterion itself may be inadequate or biased, as this study intends to explore with respect to the MEA. "The ultimate criterion can best be described as a psychological construct. Thus, the process of determining the relevance of the immediate [the test item and distractors at issue] to the ultimate [the underlying objective being measured] criterion becomes one of construct validation. That is, the assessment of the relevance of our measures...involves

---

[14] Linn, 205.

[15] Jensen, 375.

determining the 'meaning of the measurement.'" [16] Jensen argues that:

> [a] predictor is biased if it either overestimates or underestimates an individual's criterion performance depending on his group membership. A predictor is biased if it correlates more with group membership than with the criterion it is intended to predict, for under this condition [the use of test scores in qualitatively comparing the effectiveness of school[s] reward[s] or penalize[s] on the basis of...group membership rather than just on the basis of those individual traits that are in fact relevant to the criterion. [17]

RESEARCH QUESTION AND HYPOTHESIS

Research Question

Does Maine's French bilingual/English fluent language pupil group perform unlike its monolingual English pupil group, across all the ability intervals, on some of the one hundred common reading and mathematics items in the eighth grade edition of the 1989, Maine Educational Assessment Instrument?

Hypotheses

The difference in the probabilities of item level success on the 1989, common MEA mathematics items between the two subgroups: the French bilingual/English fluent pupil group, and the monolingual English pupil group is equal for all ability intervals.

---

[16] M.J. Kavanagh, A.C. MacKinney, and L. Wolins, "Issues in managerial performance: Multitrait-multimethod analysis of ratings," Psychological Bulletin, 75, (1971): 35.

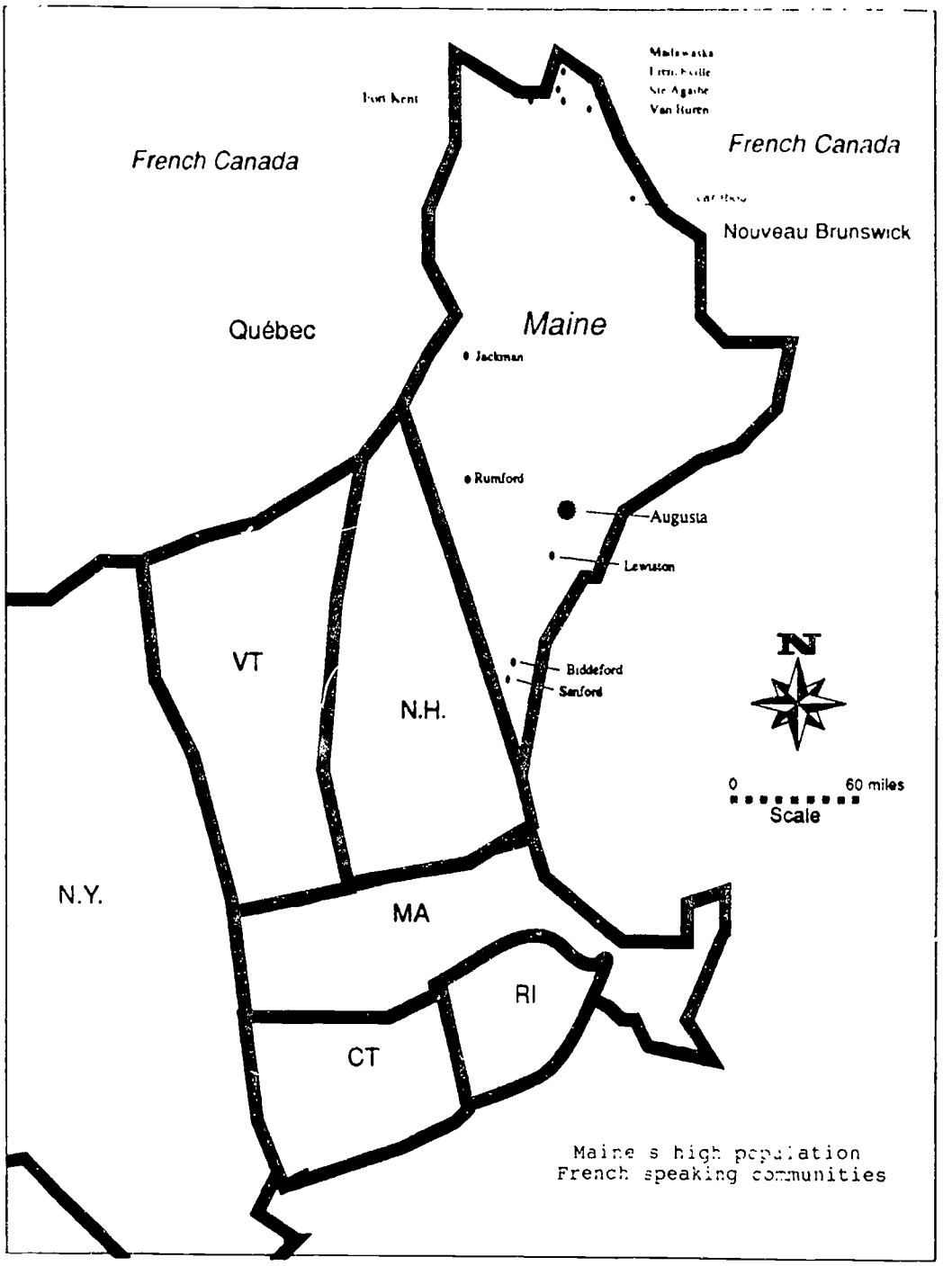
[17] Jensen, 48.

The difference in the probabilities of item level success on the 1989, common MEA reading items between the two subgroups: the French bilingual/English fluent pupil group, and the monolingual English pupil group is equal for all ability intervals.

Null Hypotheses

There will be no difference in the probabilities of item level success on the 1989, common MEA mathematics items between the two subgroups: the French bilingual/English fluent pupil group, and the monolingual English pupil group.

There will be no difference in the probabilities of item level success on the 1989, common MEA reading items between the two subgroups: the French bilingual/English fluent pupil group, and the monolingual English pupil group.


French Canada
French Canada
Nouveau Brunswick
Québec
Maine
Jackman
Rumford
Augusta
Lewiston
Biddeford
Sanford
VT
N.H.
N.Y.
MA
RI
CT
N
0
60 miles
Scale
Maine s high population
French speaking communities

CHAPTER TWO

LITERATURE REVIEW OF PRIOR STUDIES

There are many significant factors affecting the validity of assessments across all subject areas of bilingual children. Examples of these issues include:

1. determining which language is dominant and in which language the child should be tested

2. recognizing how bilingual children may use language in ways that are qualitatively different from that of monolingual children

3. recognizing the influence of cultural differences and the local environment on conditions of assessment

4. overcoming insufficiencies of existing diagnostic instruments such as lack of consideration of cultural and linguistic factors that could affect norming. [18]

In conducting this literature review, an attempt was made to look for indications in prior studies under each of the four areas outlined above.

DOMINANT LANGUAGE

This section of the literature review looked at studies emphasizing the importance of determining the pupil's dominant language and by extension, the language in which the pupil ought to be tested. Clearly if the pupil's

---

[18] "Assessing the Language Difficulties of Hispanic Bilingual Students" (ERIC Clearinghouse on Handicapped and Gifted Children, Reston, Virginia, August 1989), Abstract 23, ERIC, ED 321427.

dominant language context is French, for example, it is inappropriate to administer a test in the English language unless the pupil is demonstrably bilingual/English fluent. For unless the pupil meets all of the criteria specified in Appendix B for bilingual/English fluency, s/he will certainly be placed at a disadvantage if tested in other than in the dominant language.

The problem inherent in the above discussion of language dominance is of course inadequate or supperficial determination of the fluency status. Uninformed educators responsible for certifying the language status of pupils for test-taking purposes are certainly having a "disabling" impact on certain pupils inappropriately or improperly designated bilingual/ English fluent. Cummins provides an historical perspective on the use of the English language for educational assessment of English deficient (minority language) pupils:

> Historically, assessment has played the role of legitimizing the disabling of minority students. In some cases assessment itself may play the primary role, but more often it has been used to locate the "problem" within the minority student, thereby screening from critical scrutiny the subtractive nature of the school program, the exclusionary orientation of teachers towards minority communities, and transmission models of teaching that inhibit students from active participation in learning....From the present perspective, however, it must be emphasized that [biased] assessment is carried out by well-intentioned individuals who, rather than challenging a socioeducational system that tends to disable minority students, have accepted a role definition and an educational structure that makes

[biased] assessment virtually inevitable. [19]

Dolson indicates that... "[n]ationally, it has been estimated that at least 3.4 million school-children are limited in the English language skills needed to succeed in school programs designed for native speakers of English."[20] That is to say, educators responsible for testing decisions affecting cross-cultural pupils must be cognizant of the possiblity that some of their clients may be adversly affected by testing unless precautions are taken to insure that language dominance has been taken into consideration and accommodated for.

The State of Connecticut Department of Education offers the following administrative guideline on the assessment of questionable bilinguals, that is, children whose second (English) language skills are unknown in terms of their competency for being tested in English:

> In order to determine the language that will be used for testing and instruction, the dominant language of the student must first be established through comprehensive language study. A description of dominance, rather than a report of test scores, is more effective for further evaluation as well as for language planning. [21]

---

[19] Jim Cummins, "Empowering Minority Students: A Framework for Intervention," Harvard Educational Review, 56, no. 1 (February 1986): 28.

[20] Dolson, 16.

[21] Tom B. Gillung, ed. Providing Special Education Services to Limited-English-Proficient Handicapped Students, Administrative Guidelines, (Connecticut: Department of Education, 1990), 21.

It is clear that children exploit their facility for informal language acquisition. The use of street language facility, however, does not support the assumption that such children should be able to take an objective-referenced test in English. It must be clearly established for example, that the French-speaking children in Caribou, Maine, are truly bilingual/English fluent speakers based on such clearly defensible guidelines on assessment as those provided educators in Connecticut:

> Step 1. Identify home language usage through home language survey or interview.
>
> Step 2. Review language dominance by checking results of assessment procedure. Remember that dominance varies significantly relative to areas and situations.
>
> Step 3. Aggregate and interpret all available data. Consider if:
>
> a. a monolingual (single language) evaluation procedure will be as effective as a bilingual (dual language) evaluation in capturing the student's strengths and weaknesses, or if
>
> b. a bilingual (usage of two languages) evaluation would be more effective than a monolingual English evaluation in capturing the student's strengths and weaknesses. [22]

Olson supports this reasoning of the Connecticut State Department:

> ...the degree to which ...linguistic knowledge is

---

[22] Ibid., 65.

> conventionalized and formalized need not be very great in oral contexts since the listener has access to a wide range of information with which to recover the speaker's intentions....To serve the requirements of written language, however, all of the information relevant to the communication of intention must be present in the text....Once this [information has been provided], children or adults have sufficient basis for constructing the meaning explicitly represented by the text. [23]

Reichman and Zyskowski also explored informal versus formal language use:

> [w]hen assessing the student's level of oral English, it is very important to differentiate between the use of language for interpersonal communication purposes and for school learning activities. While a student may appear to have a good command of oral English, a thorough assessment may indicate that the student has not mastered enough English to use the language successfully for instructional purposes. [24]

McBay notes that, "[t]oo often, [Alaskan and American Indian] Natives who have problems with English proficiency are miscategorized as special education students, producing in them a sense of inferiority." [25]

These studies show that informal, subjective methods of determining language proficiency with respect to test-taking

---

[23] David R. Olson, "From Utterance to Text: The Bias of Language in Speech and Writing," Harvard Educational Review, 47, no. 3 (August 1977): 277.

[24] Susan Reichman, and Gloria Zyskowski, "Testing Approaches and Uses with Bilingual, Special Needs Students" Paper presented at the Annual Convention of the Council for Exceptional Children, 28 March-1 April 1988, ERIC ED 300 944.

[25] Shirley M. McBay, ed. Education That Works: Quality Education For Minorities Project (Cambridge, Massachusetts: Massachusetts Institute Of Technology, January, 1990), 25.

competencies are at best suspect. They demonstrate that accurate determination of the dominant language is pivotal in righting some of the wrongs generated by differential validity in standardized testing. This point will be further developed later in this chapter with specific reference as to how a pupil's English fluency is determined before taking the MEA. It should be understood that a pupil's social language skills as used in peer discourse are very different from the higher order language skills required for reasoning on a testing instrument. Since this study is looking for items indicating differential validity between pupils of identifiably equal abilities, equal language ability in terms of English language fluency should be a fundamental prerequisite for test taking.

LANGUAGE USE

This second section of the literature review looked at studies that considered how bilingual children use language in qualitatively different ways from monolingual children.

For example, Miller-Jones has shown that accurate assessment of cognitive abilities in language minority individuals is encumbered by several factors. Chief among these is the difficulty of inferring underlying cognitive processes from performances on standardized tests. Miller-Jones asserts:

> Recent developments in contexturalist analyses of cognitive performance, such as cultural practice theory, argue

> (a) that skills are acquired in specific learning activity contexts and therefore tests of generalized cognitive functioning will inevitably provide a less than accurate portrayal of individuals' capacities; and
>
> (b) that appropriate assessment requires an understanding of the constraints that govern access to a person's knowledge and regulate that deployment of concepts and reasoning processes.[26]

The implications of Miller-Jones research to a study in item bias analysis are enormous. The reading section of the MEA studied in this research was driven by three reading process objectives: general comprehension, the use of management strategies, and the use of reference skills, through the analysis of different passage types requiring an understanding of passage content and practical use of passage material. Both long and short passages were used.

Additionally, the mathematics section of the MEA was driven by three process categories: procedural components, concept and knowledge elements, and problem solving segments. Also included were five content categories requiring of the pupil knowledge of numbers and numeration, awareness of variables and relations, cognizance of geometry, understanding of measurement, and proficiency in problem-solving skills.

If, as Miller-Jones suggests, the appropriate assessment of language minority pupils assumes an

---

[26] Dalton Miller-Jones, "Culture and Testing," American Psychologist, 44, no. 2 (February 1989): 360-366.

understanding of and provision for assessment strategies designed to recover testing information acquired in specific learning activity contexts; and furthermore, requires an understanding of the constraints that govern access to a person's knowledge and regulate that deployment of concepts and reasoning processes, then item bias is almost certainly going to occur if tests are developed without taking into account the learning activity contexts of the pupils targeted for testing. Also, test developers must be cognizant of the clients' culturally based reasoning processes and constraints on the clients' abilities to access knowledge and to deploy such concepts as the test might demand. Provisions for accommodating cultural (language minority) differences with an eye to preventing the types of culturally driven knowledge access problems that are suggested by Miller-Jones at the time the test is constructed must become axiomatic in test development for instruments that will be used in a cross-cultural contest. Having said this, however, will certainly not simplify the psychometrician's task of eliminating item bias. For discovering the pupils' learning activity contexts and the concomitant keys to accessing knowledge through understanding the reasoning processes from each socio-linguistic subgroup in our vastly complicated cultural polyglot may just be more than one can reasonably expect from test developers designing tests for a multicultural

context.

Henry and Pepper in their study on Indian learning styles conclude a section of their study with some thoughts which confirm the Indian elders' concerns:

> ...[T]here is a growing body of research to suggest that distinctively different practices, one stressing observational learning and others emphasizing learning through verbalization, have fostered the development of different styles of learning among Indian and European-American children. Many European-American children, by virtue of their upbringing and their linguistic exposure, are oriented towards using language as a vehicle for learning. Indian children have developed learning styles characterized by observation and imitation, a culturally based difference with significant educational implications. [27]

By extension, unique learning styles producing unique pupil/teacher relationships, probably require unique assessment strategies. At least consideration must be given to the possibility that test format and administration procedures may create test biases, with respect to these very different cultural groups tested for their academic performance.

Miller-Jones ends his discussion by concluding that the "[c]ultural practice theory has moved us away from a conception of cognitive skills as stable trait-like personal qualities and offered in its place the idea that

---

[27] Steven L. Henry and Floy C. Pepper, "Cognitive, Social, and Cultural Effects on Indian Learning Style: Classroom Implications," The Journal of Educational Issues of Language Minority Students, 7, Special Issue (Summer 1990): 90-91.

competencies develop in the context of culturally based activities or practices." [28]

Cummins raises similar issues. His research indicates:

> ...support [for] the hypothesis that bilingualism promotes an analytic orientation to both linguistic and perceptual structures[;]...that bilinguals are more semantically oriented than unilinguals and have a greater awareness of certain properties of language is equivocal[;] ...that bilinguals had longer response latencies on a word association task than unilinguals....A major difficulty in interpreting these studies...is that the measures used to assess metalinguistic skills usually only have face validity. Where correlations between tasks are reported...they tend to be low, thus raising not only the empirical validity question but also the theoretical question of what the dimensions are of the construct of metalinguistic awareness or skill and what developmental stages it goes through.[29]

Reynolds also comments on a related element with respect to the Immersion Model in which pupils are immersed in studies entirely in their own mother tongue (L1):

> ...overall findings from standardized testing of achievement in English language arts reveal that immersion students perform on par with their monolingual counterparts despite receiving the bulk of their schooling in a foreign language....The results from controlled comparison studies in both Canada and the United States consistently indicate that immersion students do as well as or better than their monolingual peers in the subjects tested. [30]

---

[28] Miller-Jones 360-366.

[29] Peter Hormel, ed. Childhood Bilingualism: Aspects of Linguistic, Cognitive, and Social Development, (New Jersey: Lawrence Erlbaum, Assoc., 1987), 70-71.

[30] Allan G. Reynolds, The Cognitive Consequences of Bilingualism, "ERIC Clearinghouse on Languages and Linguistics News Bulletin," (Washington, D.C.: Volume 14, Number 2, March 1991), 6.

Kessler and Quinn likewise argue that additive bilinguals (that is to say, pupils whose academic studies in English (L2) have not subtracted from their development in their own mother tongue L1) significantly ($p<.001$) outperformed the monolingual group on both the hypothesis writing measure and on the syntactic complexity measure in their hypothesis writing experiment:

> Results of this study of additive bilingualism indicate that bilingual children proficient in both their languages in the sense that they use each and are literate in each as a result of participation in a bilingual education program outperform monolingual peers when given the same instruction by the same teacher in formulating scientific hypothesis. Not only do they exhibit superior performance on tasks which require aspects of divergent thinking but also they manifest a level of syntactic complexity in their second language higher than that of the monolinguals in expressing their solutions to science problems. [31]

CULTURAL AND ENVIRONMENTAL INFLUENCE

The third section of the literature review looked at studies that considered how the influence of cultural differences and the local environment influence pupils' test performance.

Cultural differences and the local environment influence how pupils think and look at the world. The content of test items and the processes required for their execution are inexorably affected by the pupils' cultural

---

[31] C. Kessler and M. Quinn, "Bilingualism and Science Problem Solving Ability," Bilingual Education Paper Series, Vol. 4, No. 1, 1980, Dissemination and Assessment Center, California State University, Los Angeles, 72.

and linguistic perspective on life. Test developers and those who develop norming standards must plan for the eventuality that results may be called into question because testing across cultures is a complicated and inexact science: one that must be looked at more for trends and indicators, than for proof of cognitive growth and development based on the school's curriculum and teachers.

The studies reviewed so far in this section indicate that cultural differences and the local environment make generalizing from test results a highly problematic operation where multicultural subgroups are concerned.

Rhodes' studies indicate that Native American populations score farther from the norm on standardized tests than does any other minority population. He suggests that their perception of reality or world view contributes to this significant difference. His conclusion:

> It is becoming more apparent that the use of standardized tests for assessment of minority populations is questionable, both from a content and a process viewpoint. Therefore, it is essential that such tests, if they are used at all, are used in a sensitive and appropriate manner which takes into consideration that they may be inaccurate and may give absolutely misleading information concerning the students to whom they are administered. [32]

The following review of the hearings held by the

---

[32] Robert W. Rhodes, "Standardized Testing of Minority Students: Navajo and Hopi Examples," Paper presented at the Annual Meeting of the National Council of Teachers of English, 78th, St. Louis, MO, 18-23 November 1988, ERIC, ED 299 587.

National Commission on Testing and Public Policy and the National Association for Asian and Pacific American (APA) Education in November 1988, raised a number of issues that are germane to the current study with respect to the influence of culture and local environment on pupils' test performance. Twelve recommendations emerged from these hearings, but only those with direct ramifications to possible testing bias as a result of cultural and environmental influence on pupil test performance are presented here:

> o When comparing standardized test scores across school systems...adjustments to test scores should be based not only upon the percentages of black and Hispanic students in a state, but also the percentage of APA's [or, upon the percentages of all language minority groups in the state.]
>
> o Nearly all of the testing issues of concern to APA's [language minority pupils] are related to differences in the linguistic and sociocultural backgrounds of APA's and the majority population.
>
> o Generalization of testing policies and testing research findings from one subgroup of APA's [language minority groups] to another with differing configuration of linguistic and sociocultural characteristics should be made with extreme caution.
>
> o It should...be emphasized that, while biased assessments of APA's [language minority groups] may be inadvertent, due to lack of knowledge and understanding of their linguistic and sociocultural characteristics, the effect of inadvertent bias is the same as deliberate bias. APA [language minority group] access to equal educational and economic opportunities is unfairly

limited. [33]

Another Indian study by Florey and Tafoya pointed out factors contributing to the inappropriateness of standardized assessment tools for use with Indian pupils without regard to cultural and local environmental factors affecting pupil test performance:

> ...neglect of subcultural values, abilities and knowledge in assessment instruments and procedures; use of exclusive training [of examiners] in application of middle class measurement instruments; belief that object measurement is the only way to conduct assessment; inadequate attention to problems of motivation and negative reactions to the examiner; failure to include sufficient numbers of minority students in standardization calculations; and lack of knowledge about culturally valued talents of American Indian students. [34]

The cultural implications in the foregoing citation, if ignored by testing experts, will result in untrustworthy results if such tests are to be administered to these subpopulations and if comparisons are to be drawn with the majority population. The cultural (mis)use of time-on-task in testing, lack of trust for non-subgroup examiners, value conflicts with domains being tested, motivational problems,

---

[33] Tony C.M. Lam, "Testing, Opportunity Allocation, and Asian Americans," The Proceedings of a Hearing Co-Sponsored by the National Commission on Testing and Public Policy and the National Association for Asian and Pacific American Education, Honolulu, Hawaii, 11 April 1987, 15-17, ERIC 297 058. A report.

[34] Janice Florey and Nancy Tafoya, "Identifying Gifted and Talented American Indian Students: An Overview," ERIC Digest, EDO-86-0024, (March 1988): 17-14, ERIC ED 296 810.

and small numbers of pupils being tested can all lead to cases where the subpopulation experiences test performance differences which are likely to include item differential validity in selected instances.

Further analysis of literature exploring the impact of cultural and linguistic effects on standardized assessment found that Mishra, using the likelihood ratio chi-square statistic analyzed by the log-linear technique for analyzing multi-dimensional categorical data in the WISC-R, found support for the "notion that the experiential backgrounds of culturally different subjects may not have provided opportunities to learn the content of certain vocabulary items...contrary to the notion that the vocabulary items in the WISC-R allow 'any recognized meaning of the word disregarding any elegance of expression.'[35] " [36]

Mishra's discussion indicates:

> The performance of Anglo, Mexican-American, and Navajo subjects on verbal items of the WISC-R [demonstrated that] 15 items provided an indication of non-homogeneous performance [differential validity] by subjects from three different cultural backgrounds. These items are from the Information, Similarities, the Vocabulary subtests. Such a discrepancy in performance of three groups of subjects on some of the items of three subtests detected to be biased in this study

---

[35] D. Wechsler, Wechsler Intelligence Scale for Children-Revised, (New York: The Psychological Corporation, 1974), 161.

[36] Shitala P. Mishra, "Ethnic Group Bias in WISC-R Verbal Items," Paper presented at the Annual Meeting of the American Psychological Association, Anaheim, California, 26-30 August, 1983, ERIC ED 236 172, 7-14.

are consistent with the findings of recent investigations in this area.[37] [38]

Questions raised by these studies pose challenges for those looking to standardized testing as the answer to America's educational accountability problems. The United States today is too complex in a multicultural sense to make assessment the easy answer. Careful analysis of test items, and pretesting of groups involved, producing local norms, may be required if credibility is to be (re)established in communities that may have been subject to alleged inappropriate testing, and inappropriate use of problematic test results as indicated in these studies.

Lucas, Henze, and Donato make a statement with respect to Latino pupils that is germane to the current study:

> ...the diversity among students cannot simply be ignored. While the schools recognized the importance of integrating language-minority students with mainstream students and of providing equally challenging instruction for all students, they did not try to minimize differences among mainstream and Latino students or among Latino students themselves. Approaches to schooling that value linguistic and cultural diversity and that promote cultural pluralism were welcomed and explored whenever possible. Students' languages and cultures were incorporated into school

---

[37] Mishra refers readers to: Cotter, D.F. & Berk, R.A.. Item bias in the WISC-R using black, white, and hispanic learning disabled children. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April, 1981.

[38] Ibid., Mishra 13.

programs [furthering] academic success. [39]

Such an approach to working with language minority pupils is very much in keeping with the issues underlying this section: cultural and local environmental issues that can affect test performance. Not to minimize legitimate cultural differences, and not to overlook the need to adapt testing instruments and procedures to better fit the subgroups in question can improve testing validity and reduce instances of differential validity.

OVERCOMING DIAGNOSTIC DEFICIENCIES

The final section of the literature review looked at studies that considered how to overcome deficiencies in existing diagnostic instruments.

Lupi and Woo explored testing bias in the assessment of handicapped and limited English-proficient students of East Asian origin. This study demonstrates how the dominant culture, by not understanding the complex cultural differences found in such divergent civilizations as China, Hong Kong, Japan, Korea, and Vietnam, develops assessment instruments containing cultural biases and linguistic distortions. Domains in testing using objects and concepts in the students' culture, e.g., spirits like the elf would be incomprehensible to East Asian pupils whose supernatural

---

[39] Tamara Lucas, Rosemary Henze, and Ruben Donato, "Promoting the Success of Latino Language-Minority Students: An Exploratory Study of Six High Schools," Harvard Educational Review, 60, no. 3 (August 1990): 337.

beings are quite different; or, the use of objects in tests requiring experience with different degrees of familiarity, e.g.: What animal does bacon come from? Whereas bacon is largely unknown in China, ham would be a better choice. Other examples are facts not taught in the child's home country; or, different customs; or, different ethics; or, unfamiliar or culturally inappropriate forms of assessment; or, linguistic distortions; or, different language structures; or, untranslatable terms; or, terms with different levels of difficulty; or, different levels of discourse. [40] They indicate how biases may be included in standardized tests, and offer explanations and suggestions for modifying or eliminating them. Primarily they recommend that culturally and linguistically appropriate instruments for East Asian pupils must be developed and validated, and until this is accomplished, evaluators must be trained to recognize that commonly used assessment instruments are likely to be biased requiring alternative scoring or administration procedures.[41] Surely such suggestions by extension are good advice to evaluators assessing pupils of all subgroups, such as francophones or Native Americans of Maine, and not just East Asian pupils.

Miller-Jones in "Culture and Testing" makes

---

[40] Marsha H. Lupi, and Joseph Yam Ting Woo, "Issues in the Assessment of East Asian Handicapped Students," Diagnostique, 14, no. 3 (Spring 1989): 151-4.

[41] Ibid., 158-9.

recommendations for assessment based on his research summarized earlier. He indicates that "[i]t is important to consider children's repertoires of cognitive process and the contingencies affecting their use in order to improve success in assessing the competencies of children from diverse backgrounds." For testing he suggests this means:

> 1. For any knowledge domain one must specify the possible processes that may be involved or elicited by tasks and stimuli.... In this, cross-cultural psychology agrees with the approach typically associated with information-processing methods of task analysis....
>
> 2. One should use multiple tasks, with a variety of different materials, with the same individual or population, and not assume generality from a single measure. Furthermore, one has to demonstrate that 'the range of tasks used to sample a hypothetical domain of intellectual activity actually covers the domain in a representative manner' (LCHC, 1982,p.654).
>
> _____Laboratory of Comparative Human Cognition. (1982). Culture and intelligence. In R.J. Sternberg (Ed.). Handbook of human intelligence (pp. 642-722). New York: Cambridge University Press.
>
> 3. It is critical that 'the tasks used to sample the domain in question do so for the culture in question' (LCHC,1982,p.654).
>
> _____ Ibid., 654.
>
> 4. Validation procedures need to better establish the relationship between cognitive operations tested and the acquisition of school concepts and skills such as reading, mathematics, writing, and science.
>
> 5. Because representations of knowledge may be configured and accessed differently by individuals varying in cultural background, it is important to develop assessment procedures that permit and direct examiners to probe for the reasoning behind a child's response to an item. It is often not a failure to use a self-generated cognitive strategy that accounts for poor test performance. Rather it is the inconsistent application of an approach or

the failure to recognize the cognitive operation the task calls for. [42]

Miller-Jones' recommendations are demanding from a psychometric standpoint requiring a radical departure from the status quo in educational assessment in general. Given the current and ever growing multicultural impact on this society, however, a national initiative to analyze his advice as well as to map out a plan of action is essential. These points will be developed more fully in Chapter V.

Recommendations that grew out of the Florey and Tafoya study urged assessment specialists to:

> o...consider whether the child exhibits outstanding powers in one or more abilities valued by the child's culture,
>
> o...measure at a bright average level in national norms in both ability and achievement,
>
> o...[consider whether the child] demonstrates creativity, and shows leadership potential,
>
> o...access [both] verbal and nonverbal responses,
>
> o...provide adequate time for students to answer,
>
> o...develop questioning procedures to elicit multiple responses on items giving credit for such responses,
>
> o...assess a wide range of abilities,
>
> o...and use a matrix rather than one factor for making decisions. [43]

Giles argues that nonspecialized educators are at a

---

[42] Miller-Jones, 364-5.

[43] Florey, 14-17.

disadvantage from the outset, rendering them of questionable benefit to the needs of Indian pupils. "There are culturally distinctive aspects of the Native American society which make it difficult at best for teachers to effectively deal with Native American students without specialized training." [44] And by extension, Giles' conclusions are true for nonspecialized educators working with any language minority pupil to a greater or lesser extent.

Likewise, Delpit in a study of African Americans suggests:

> ...that appropriate education for [language minority children] can only be devised in consultation with adults who share their culture...[and who are] allowed to participate fully in the discussion of what kind of instruction is in their children's best interest. ...[T]hose who are most skillful at educating [these] children do not allow themselves to be placed in 'skills' or 'process' boxes. They understand the need for both approaches.... I contend that it is those with the most power, those in the majority, who must take the greater responsibility for initiating [this consultative] process. [45]

Kleinfeld, in looking for ethnic bias in Alaska's Statewide ITBS (Iowa Test Of Basic Skills) Program discovered the Riverside Publishing Company, which publishes

---

[44] Karen Knoland Giles, Indian High School Dropout: A Perspective. (Milwaukee, Wisconsin: Midwest National Origin Desegregation Assistance Center, University of Wisconsin-Milwaukee, 1985), 27.

[45] Lisa D. Delpit, "The Silenced Dialogue: Power and Pedagogy in Educating Other People's Children," Harvard Educational Review, 58, no. 3 (August 1988): 296-7.

the ITBS

> ...attempts to minimize such bias in a number of ways. Every test item undergoes three different expert reviews for possible bias, including a review by a panel of independent experts selected on the basis of geographic region and ethnic composition. Quantitative analyses of test results from an item tryout study are also done to see if minority students get lower test scores than non-minority students of similar abilities. [46]

Likewise, Wolfram, noting the fact that lower class and minority group test takers consistently score lower on standardized tests than their middle class Anglo counterparts, recommends testing knowledge he believes educators should have in order to be fair to test takers who speak vernacular dialects of English:

> 1. Consider what the test claims to be measuring in relation to what it actually measures;
>
> 2. Consider what assumptions about language underlie the test;
>
> 3. Consider what kinds of language-related tasks are necessary for the test taker to participate adequately in the test;
>
> 4. Examine demographic information provided in the test manual about linguistic and cultural groups on which the test was standardized;
>
> 5. Consider how test results can be interpreted for different dialect groups.
>
> For language specialists (e.g., speech and language pathologists, language arts educators), the following additional recommendations should be

---

[46] Judith Kleinfeld, and others, "The Alaska Statewide Student Testing Program: Are The Tests Biased?," Alaska Educational Research Association, Anchorage, Alaska, April 1991, 9.

considered:

1. Become familiar with the linguistic characteristics of communities represented by test takers;

2. Be able to identify linguistic responses to test questions that might be attributable to dialectic differences;

3. Complement standardized, formal measures of language with assessment strategies more focused on underlying language ability in real communicative contexts;

4. Gather ethnographic information on the language use of test takers from non-mainstream communities in a natural setting. [47]

In a related study, Cabello indicates three possible sources of bias in the dual language (English/Spanish) versions of the CTBS (California Test of Basic Skills): 1. problems inherent in the translation; 2. the match between the test and instructional material; 3. intervening cultural values. [48]

If decisions are to be made on the basis of standardized achievement tests where bilingual education is a factor, educators must be aware of a number of areas of concern:

1. All of the biased test items in CTBS...indicated the writer's assumption about the

---

[47] Walt Wolfram, "Dialect Differences and Testing" (ERIC Clearinghouse on Language and Linguistics Center for Applied Linguistics, Washington, D.C., October, 1990), ERIC, ED 323 813.

[48] Beverly Cabello, "Potential Sources of Bias in Dual Language Achievement Tests," Center for the Study of Evaluation, California University, Los Angeles, California, 1981, 20.

intended audience; that is, that the Spanish language reader would perceive the same implied values from the passage as would the English-speaking reader. The fault lies not with the **passage** but with the **question** because it elicits knowledge which is **external** to the passage and varies from culture to culture. Thus, culture interferes here not with superficial features of the item but with the assumptions underlying the test question.

2. The CTBS and its Spanish version are, for the most part, equivalent in terms of vocabulary, content, and format....However, examination of curricular match in terms of vocabulary and general topics suggests that the English language version has a stronger match to English basal readers. Since the content and vocabulary are the same across language versions, the tests' closer match to English basal readers may reflect the fact that monolingual English and bilingual program curricula are probably different in terms of vocabulary and content....The monolingual English curricula assume that all students are proficient in English and thus they concentrate instruction on basic skills or other areas. The bilingual curricula focus instruction on the teaching of English as a second language, teaching other subject areas using a combination of the primary language and English.

3. The problem of cultural interference [that is] ...divergent interpretation of a passage (by two cultural groups) caused by the interjection of one group's cultural attributes into one or more features of the passage ...suggests that there may be subtler, more elusive forms of cultural bias such as the interjection of values or associations which reflect one culture and not another. [49]

By extension, the collective impact of the foregoing recommendations and observations must be extended to play a role in the development of assessment instruments like the MEA for use with Maine's cultural/linguistic (language

---

[49] Ibid., 21-23.

minority) groups. Their relevance must not be overlooked. This assertion is especially notable, considering that Maine's MEA is "checked for item bias" only by way of superficial analysis by the classroom teachers who develop it. It is certain that these teachers are ill-prepared for doing a qualified job with such an analysis within the context of the research cited in this section. It should be noted that the MEA's developers do check item performance through item characteristic analysis after the items have been pre-tested in the matrix sampling questions, and before they are used in the common questions. There is not, however, any form of cross-cultural item analysis done by the developers.

Appendix B presents a survey sent by the Maine Department of Education to ninety one administrators of schools enrolling at least 20 minority language pupils as indicated by responses to the annual Lau home language survey conducted by the Maine Department of Education in 1990. Sixty-three (69%) of these school administrator surveys were completed and returned. The administrators were asked to explain how they had determined if a pupil was competent to take a test in English; i.e., if the pupil was bilingual/English fluent. The actual question asked was: "In selecting the MEA subgroup response 'bilingual/English fluent,' the following determinants were used:."
The options available for choice were as follows:

_____ The results of language assessment instrument (e.g., Language Assessment Scales, Language Assessment Battery, Individual Proficiency Test, Macaulitis...)
_____ The results of an intelligence test
_____ The results of a standardized achievement test
_____ Academic grades
_____ Conversational English proficiency through subjective observation
_____ Writing competency from in-class observation
_____ Reading competency from in-class observation
_____ The results of a language assessment committee's documentation
_____ Student self-identification as "bilingual/English fluent"
_____ Home language survey indication of bilingualism
_____ The subjective judgment of another staff person(s)
_____ The subjective judgment of the parent(s)
_____ Other:_____
_____ I do not know how the determination was made.

Fully forty of the sixty-three administrators (63%) reported using: "Conversational English proficiency through subjective observation," to establish language competency. Clearly such an informal, subjective approach could prove disastrous for some children who might seem to be more proficient than they actually are.

The relationship between the relative language proficiencies of bilingual children and their scores on tests administered in one or both languages was studied for students referred to special education because of suspected learning disabilities. [50] Bilingual Hispanic students who

[50] Cheryl Yelich Wilkinson and Wayne H. Holtzman, Jr., "Relationships among Language Proficiency, Language of Test Administration and Special Education Eligibility for Bilingual Hispanic Students with Suspected Learning Disabilities," Paper presented at the Annual Meeting of the American Educational Research Association, 5-9 April 1988, TM 012 611, ERIC ED 301 604.

were considered as limited English proficient were assessed using intelligence and achievement tests in English and Spanish. The effect of the language of test administration on IQ scores was difficult to assess; however, between 9% and 17% of the group tested qualified for learning-disabled services on the basis of English (but not Spanish) scores. This finding would appear to underscore the need to consider native language assessment for all bilingual children.

Maine has no provision for carrying out the MEA in a language other than English, nor does it have a trustworthy method for determining if a pupil is sufficiently competent in English to take the test, as was shown by the survey provided in Appendix B.

In short, the literature suggests that assessment of bilingual/English fluent pupils can be problematical in that test items can generate differential validity in the scores of minority language pupils. In fact, Cummins believes that assessment has been used to place language minority pupils at a disadvantage, thereby focusing attention away from "subtractive" school programs, and teachers who would prefer to exclude such pupils from the mainstream. Such use of testing, locates the "problem" within the language minority student, and promotes a model of teaching that inhibits pupils from accessing an active role in the learning process.

Educational reform initiatives of the past decade have

jumped at assessment as the solution to school accountability. But predicating reform strategies on a single strategy like assessment (without taking into consideration the complex and rapidly changing demographics impacting this society) has forced the issues considered in the current study to the surface in Maine.

Ferdman has concluded that cultural diversity is inextricably tied up in the relationship between literacy and the individual, and that if our society is to focus on extending literacy, it must do so within light of our cultural pluralism.

Not to recognize and accommodate for cultural diversity across the multicultural spectrum, while at the same time demanding educational accountability verified through various assessment strategies, will keep testing bias (differential validity) in the mainstream of discussion in Maine. This testing bias issue has been well documented in the Black and Hispanic educational realm, but sub-populations like those looked at in this literature review have not received adequate consideration. Jensen has clearly shown that error of measurement (unreliability) and error of prediction (invalidity) are directly related to an individual's group membership.

AFTERWORD.

A recent article in "U.S.News & World Report" concludes with a caveat which is apropos as a final thought to the

current literature review:

> The lesson of the corruption of standardized testing in recent years seems to be that the nation mistakenly has tried to ratchet up accountability in public education on the cheap. Multiple-choice tests of low-level skills may be relatively inexpensive to administer, particularly in the absence of tough test security. But as they are used today, many are of dubious educational value. This should serve as a warning to the advocates of national testing. If the nation is to build a new national examination system..., it needs to invest the resources necessary to build tests with high standards and rigorous security [which are multiculturally unbiased]. If it doesn't, the testing debacle in America's schools may only get worse. [51]

---

[51] Thomas Toch and Betsy Wagner, "Schools for Scandal," (Washington, D.C.: U.S. News & World Report, Inc., Volume 112, Number 16, April 27, 1992), 72.

CHAPTER THREE

DESIGN OF THE STUDY

RESEARCH DESIGN AND POPULATION

This study analyzed part of the eighth grade test data gathered in the State of Maine, through the administration of the Maine Educational Assessment (MEA) instrument between 24 October and 4 November 1988. The monolingual English sample data included 1302 randomly selected pupil records, approximately 7% of the total state-wide monolingual English speaking population, from which 336 pupil records were matched with an equal number of French bilingual/English fluent pupils based on identical overall test scores. The 1302 pupil records was the number needed to generate the 336 matching overall (identical) test scores with the French pupils on overall test performance.

The 336 French bilingual/English pupil records were from the communities of St. Agatha, Madawaska, Lewiston, Ft. Kent, Van Buren, and Caribou. These 336 records constituted the total population of 8th grade pupils in these six communities who were designated bilingual/English fluent. A student was considered bilingual/English fluent if there was evidence that a language other than English was used in the student's home environment and the student was proficient in all English communicative skills areas of reading, writing, speaking and listening commensurate with his/her monolingual English peers. The in formation that categorized the pupil

as bilingual/English fluent was obtained from one of the demographic indicators used in the test. If the bilingual/English fluent bubble on the answer sheet had been blackened, it must be assumed that all of the criteria for bilingual/English fluency had been met; however, as is noted in the analysis of the survey recorded in Appendix B, and following Table 2, this assumption is not always warranted. Nevertheless, such demographic information is contained in the MEA data base stored at Advanced Systems In Measurement and Evaluation, Inc. of Dover, New Hampshire.

As described above, the MEA test package delivered to all school systems had grids for gathering demographic information, including one on language status. The test administrator, usually the principal, indicates on the grid the category of English language fluency of each child in the grade being tested. The administrators are asked to follow the federal definition of limited English proficiency in identifying this subgroup as shown in Table 2.

TABLE 2	
Categories of English Language Fluency	
CATEGORY	DESCRIPTION OF CATEGORY
_____	A student is monolingual English if there is no evidence that a language other than English is used in the student's home environment. The Department of Education Lau home language surveys were used to determine if students were bilingual or monolingual.
_____	A student is **bilingual/English fluent** if there **is** evidence that a language other than English is used in the student's home environment **and** the student is proficient in **all** English communicative skills areas of reading, writing, speaking and listening commensurate with his/her monolingual English peers. The Department of Education Lau home language surveys were used to determine if students were bilingual or monolingual.
_____	A student is **bilingual/limited English proficient** if there is evidence that a language other than English is used in the student's home environment **and** the student has **limited English proficiency in one or more** of the English communicative skills areas of reading, writing, speaking, or listening.

Based on the results of a state-wide survey of elementary principals conducted by the Maine State Department of Education in the Fall of 1990, inconsistent identification patterns of language minority pupils by these test administrators are almost certainly leading to mis-identification on the language status questions of the test. [52] If this is the case, then some of the subgroups designated

[52] See appendix B.

by this MEA process are certainly not of "identifiably equal ability," and hence, do not meet the requirements of subgroups appropriate for inclusion in "a given interpretation of a test score" within the parameters of the definition of bias used in this study. It is clear that if the inconsistencies in determining language fluency for the purposes of test taking are to be eliminated, then a more trustworthy method of identification will have to be worked out. That is, the procedures for determining language fluency by the test administrators must be made more objective; it is the current level of subjectivity used in this identification process that is producing mis-identification resulting in pupils taking the MEA without the requisite English language skills.

THE INSTRUMENTS

The MEA is designed to measure student performance in grades 4, 8, and 11, in the subject areas of reading, math, writing, science, social studies, and humanities.

The tests contained a total of 929 multiple choice items, including 50 common items that all pupils take in both reading and mathematics, 20 open-ended items, and 2 essays. Table 3 provides detailed item category information.

TABLE 3	
MEA Test Item Categories	
Categories	Number of Items
Reading	217
Open-ended Reading	9
Writing Essays	2
Mathematics	196
Open-ended Mathematics	10
Science	192
Social Studies	204
Humanities	111
Open-ended Humanities	10

All test results are presented as scaled scores, because of their ease of interpretation. Scaled scores can be used to compare the relative performance of one group within a school to another, the relative performance within a school on one skill area to another, the relative performance within a school on one content area to another, and the performance of a school one year relative to its performance a previous year.

In order to develop the scale, the statewide average was set at 250 in 1985-86 for each of the five of the content areas. For these areas, scores larger than 250 mean that a pupil's performance was above the statewide average for 1985-86. For the sixth area, writing, the scale has been reset each year so that the state-wide average is 250.

Scaled scores can range from 100 to 400. If a group average is below 100 or above 400, it is printed as 100 or 400, respectively. Scores of 100 do not mean no questions were answered correctly. Such scores indicate achievement levels far below or far above the statewide average. [53]

The item internal consistency reliability coefficients [54] for 1988-89 school mean scores, computed with schools as the unit of analysis, are provided in Table 4 below.

TABLE 4

KR/20 (Kuder/Richardson) Reliability Coefficients For 1989 School Mean Scores (Internal Characteristics)

Subject	Eighth Grade
Reading	.97
Mathematics	.96
Science	.93
Social Studies	.96
Humanities	.93

Individual student scores in reading and mathematics are based on only the fifty or so common items in the subject area that all students answer. This study was based on an analysis of these common reading and mathematics items. The reliability coefficients for common items in the

[53] From explanatory materials published by the Department of Education, Division of Educational Assessment, 1989-90.

[54] Guide To The Maine Educational Assessment 1989-90; Department of Education, Division of Educational Assessment, 40.

1988-89 grade 8 tests were .87 in reading, and .90 in mathematics. [55]

To determine 1989 concurrent validity on the common items in reading and mathematics, pupil scores were correlated with scores on the ITBS, STEP, CTBS, and SRA standardized achievement tests. This standardized data was gathered by the Maine Department of Education from the respective testing companies for the purposes of fulfilling the department's need for concurrent validation statistics. These statistics are provided in Table 5 below.

TABLE 5

1989 Concurrent Validity Pupil Scores [56]

Test	8th Grade Reading	8th Grade Math
ITBS [57]	.80 n=676	.85 n=677
STEP [58]	.70 n=930	.85 n=936
CTBS [59]	.79 n=264	.80 n=265
SRA [60]	.80 n=102	NA N=150

[55] It is these individual scores that this research proposes to analyze.

[56] Guide To The Maine Educational Assessment, etc., 40.

[57] Iowa Test of Basic Skills

[58] Sequential Test of Educational Progress

[59] Comprehensive Test of Basic Skills

[60] SRA Survey of Basic Skills

Individual pupil scores on the common items in reading and mathematics were compared for the two groups described previously: monolingual English and French bilingual/English fluent pupils. The database of the Maine Educational Assessment for grade 8 for test year 1989, was available through Advanced Systems In Measurement & Evaluation, Inc., of Dover, New Hampshire. Also, the Maine Department of Education made available the common questions asked of the pupils in the eighth grade in the 1988/89 test year for use in this study.

DATA

The data for this study included 1988-1989 8th grade files retrieved from the data base of Advanced Systems. Each 1988-1989 file contained the pupil responses to the 41 Common multiple choice Reading items, the 9 Open-Ended items, the 40 Common multiple choice Mathematics items, and the 10 Open-Ended items. Open ended items typically had ten possible answers with usually three correct (acceptable) responses based on the pupil's attached reasoning.

STATISTICAL METHODOLOGY

This study replicated the statistical methodology used by H. Johnson Nenty in his study: CROSS-CULTURE BIAS ANALYSIS OF CATTELL CULTURE-FAIR INTELLIGENCE TEST. [61]

---

[61] H. Johnson Nenty, "Cross-Cultural Bias Analysis of Cattell Culture-Fair Intelligence Test," Paper presented at the Annual Meeting of the American Educational Research Association, 16-20 April 1986, TM 860 434, ERIC ED 274 668.

Nenty's pre-research analysis of testing for cross-cultural validity cited the work of C.E.Massad details of which may be found in Appendix E.

Following Nenty's approach, four item analysis strategies were used in the current study to examine cross-cultural validity: Scheuneman's modified chi-square procedure (SSX^2); Rudner and Convey's TID-45 degrees item difficulty p-value; the one-parameter latent trait Rasch model; and the Mantel-Haenszel procedure. Nenty actually used Cochran's Chi Square Test Method CTX^2 as his fourth statistical strategy, however, the newer Mantel-Haenszel procedure, "one of the most promising of the Chi-square variants," [62] was substituted for Cochran's Test Method in the current study, because of its prominent use in current cross-cultural validity studies. In his study, Nenty noted the overlap among these procedures.

DISCUSSION OF THE FOUR STATISTICAL STRATEGIES

Scheuneman's (1979) modified chi-square procedure (SSX^2) , defines an item as unbiased:

> if, for all persons of equal ability (i.e., equal total score on a test containing the item) the probability of a correct response is the same regardless of each person's cultural or ethnic group membership. With this procedure, each item is separately tested for bias by first establishing a number of ability or total score intervals on the total score on the rest. [63]

---

[62] Linn, 211.

[63] Nenty, 7.

Conversely:

> ...an item is defined as biased if individuals from different groups, who have the same total score on the test, have different probabilities of responding correctly to that item. [64]

Briefly, this is a test of independence between two categorical variables whereby, in the present case, a search is made to determine if there is a relationship between group membership and item responses on the instrument in question. This method differs from the Mantel-Haenszel procedure in that the independence of correctness of response and group is measured at ability or achievement intervals, and at each interval low, average, or high, one is looking for similar item characteristic curve (ICC) for both groups at each interval instead of summing across the whole score range as is done in the conventional Chi-square.

Another frequently used method for examining items for bias, compares item difficulty (p values) across pairs of groups. "As defined by the classical test theory: an item is biased if its p-value, or a transformation of this value, differs significantly across the groups under comparison. One version of this approach, the TID 45^0 (Rudner & Convey, 1978; Rudner et. al., 1980) entails standardizing the p-value to within-group z-value." [65]

Osterlind expands on this strategy as follows:

---

[64] Linn, 189.

[65] Nenty, 7.

> In the approach to detecting and correcting test item bias known as transformed item difficulties (TID), bias is considered to be a characteristic inherent in all test items, and the degree to which individual test items exhibit this property is the focus of attention....[S]imilar to the ANOVA strategy...relative item difficulties for two or more groups' performances on a set of test questions covering a single skill or mental construct are revealed by the interaction of groups with items...seek[ing] to identify particular biased test items by their degree of aberrance.
>
> Hence, bias as investigated with the TID approach is a relative standard; an item is considered biased when it is comparatively more difficult to answer correctly for one group than it is for the other. The assumption is that bias is indicated by a significant group difference in the relative difficulty of the item rather than by a group difference in the means, or standard deviations, of the p values or some other item difficulty index. [66]

In short, the TID approach mathematically transforms item difficulties so that the relationship between obtained difficulties for the two groups is linear. A line, in this case a 45^0 line, is fitted to points representing the difficulty of the items for each group. Distances from the line are computed for each point, confidance bands are established, and item "outliers" can be spotted easily for modification or removal from the instrument to ensure a more fair final instrument.

The third strategy used in this analysis is Rasch's, (1960) one-parameter latent trait model (ICC-1). This

---

[66] Steven J. Osterlind, "Test Item Bias," Paper series on Quantitative Applications in the Social Sciences, Sage University, Beverly Hills and London, 1983, series Number 07-030, 28-29.

strategy:

> ...stipulates a probabilistic relationship between observable item and person scores and the latent item and ability parameter assumed to underline the result of a person-by-item encounter during testing. The model holds that, given a unidimensional test, the probability of an examinee succeeding on any item is wholly determined by his/her ability on the trait that is being measured, and by the difficulty of the item. It assumes the absence of guessing and constant discrimination for all the items of the test. The model implies that 'the estimate of the item difficulty parameter will not vary significantly over different samples...[but the] item parameter will not be sample-invariant where there is culture bias which differentially affects the item probability.' Following from this 'if a test includes items which differ in cultural loadings, the special conditions required for item parameter invariance may be difficult to obtain' (Whitely & Dawis, 1974, p 175-176). To test for bias with this method, one procedure is to calibrate the items separately for each cultural group, and then compare, for each item, the Rasch difficulty parameter estimates across groups (Nenty and Dinero, 1981). [67]

Briefly this method looks for covariation (differential responses) among observed measures (subgroup means) based on latent (unobserved cross-cultural) explanations for these observed indicators. A characteristic curve is plotted for each item analyzed for each of the subgroup's item mean responses with a verticle axis indicating item difficulty, and an horizontal axis indicating ability. If differential validity is not a factor one would expect the item characteristic curves to be virtually identical.

The fourth strategy used in this analysis was the

---

[67] Nenty, 7-8.

Mantel-Haenszel Procedure. Cole and Moss say about this strategy:

> Yet another variant of the log-linear approaches is the use of the Mantel-Haenszel (1959) statistic. This statistic has had wide use in medical research studies examining dichotomous outcomes in differentiated groups with a blocking variable and is being applied as an indicator of differential item performance....
>
> The major distinguishing feature of the Mantel-Haenszel procedure is that, instead of testing against a general alternative hypothesis of any difference in correct response rates between groups, this statistic tests against a particular alternative of a common odds-ratio of correct response across all blocking (matching) categories. Practical implications of the different alternative hypotheses in test item studies are not yet known. Even so, the procedure appears to be among the most promising of the Chi-square variants. [68]

As was explained in Scheuneman above, the Chi-square method looks for independence between two variables in the relationship between group membership and item response.

TREATMENT OF THE DATA

The Statistical Analysis System (S.A.S.) software package, version 6.04, was utilized for the analyses. Each common item in the reading and mathematics test was analyzed for each subgroup using the four item bias techniques. Comparisons were made among the results provided by each analytical model to determine the degree of consistency among these four statistical strategies.

It was expected that there would be differential

---

[68] Linn, 211.

scoring between the French bilingual/English fluent pupils, and the monolingual English pupils on certain test items. When such a relationship was found, item comparisons were then made with any similar (empirical) conclusions reached by a panel of experts from the sample of items examined by them.

BIAS VALUE PARAMETERS

Initially an item was defined as biased if it was selected by all of the statistical strategies. This demarcation generated five items in both the mathematics and the reading categories. However, since eight experts had agreed to participate on the Panel of Experts described below, and since three separate subject category "expert packets" had been designed by the researcher to be analyzed by the experts, the definition of bias was slightly extended in each subject category to access a few more items for analysis.

In the case of mathematics, items 7, 23, and 31 were included in the bias definition bringing the total number of mathematics items analyzed to eight. Each of these three referenced items was selected by three of the four statistical strategies.

In the case of reading, items 18, 26, and 41 were included in the bias definitions, again because they were identified as biased by three of the four procedures, bringing the total number of reading items analyzed to nine.

Item 31 was selected by only two statistical strategies.

In each of the added items just described, the values of the Chi-squares, p-Value differences, and/or Z Scores of these additional items were established on the basis of being the next smaller adjacent values to those actually selected by all four of the statistical strategies originally defined as indicating bias. The actual bias value parameters for each of the statistical strategies ultimately included in the definition of bias are indicated in Table 6 below.

TABLE 6

Statistical Strategy
Bias Value Parameters

Strategy	Mathematics	Reading
SSX^2 Chi-square	> 4.5	> 3.3
TID-45^0 p-Value Difference	> .24	> .20
ICC-1 Z Score [69]	> 1.02	> 1.33
M-H Chi-square	> .005	> .02

THE PANEL OF EXPERTS

A sample of the biased and unbiased test items in reading and mathematics was examined by members of a panel of experts to see if they could identify those items which

[69] Negative scores indicate French group was favored by item bias.

were identified statistically as being biased. The experts were also asked to explain the reasons for the bias in the items they selected as being biased. These experts were native francophone adults closely aligned to the francophone educational experience in Maine. It was hoped for future reference, that the analysis of these experts might lead to item selection recommendations which could prevent the use of differentially valid (biased) items in subsequent testing cycles.

There were eight such experts invited to participate in this study based on a list of possible candidates submitted to this researcher by the Department of Education's Federal Projects for Language Minorities Section, Division of Administrative Services, Augusta, Maine...all accepted the invitation to participate, but only six actually carried out this commitment. The panel was drawn from an available pool of educators involved in working with multicultural pupils; of bilingual education experts; and of administrators involved in working with minority language pupils. To assure cultural and linguistic knowledge, all were members of Maine's Franco-American communities.

These experts were not asked to examine all of the items analyzed by the statistical package, because this would have entailed an unreasonably superfluous level of analysis for volunteers. Rather, they were each given a sample of the questions that had been shown to indicate item

bias through the statistical methods of selection described above, and an equal number of items randomly selected from the remainder of the item bank that had exhibited no indication of such item bias through the same processes. The panel of experts analyzed these selected questions from each category to see if they could spot the items identified by the statistical methods as exhibiting bias. The experts were asked to indicate the language, ideas, cultural differences, etc., that might explain the differences between the two groups on the items in question. By comparing before-the-fact and after-the-fact data, it was hoped that some useful insights and recommendations might be generated with respect to the reliability of expert analysis on potential item bias.

Additionally, by looking at the objectives that underlay the common questions asked to see if there were any common objectives on which the Maine French bilingual/ English fluent students indicated disparity, some additional conclusions might be drawn for the recommendations section of the dissertation.

CHAPTER FOUR

RESULTS

SUBGROUPS ANALYZED

This item bias study examined the common reading and mathematics items in the Maine Educational Assessment (MEA) test administered to Maine's eighth grade population in 1989. After noting that the French bilingual/English fluent pupils in six of Maine's predominantly French communities (Caribou, Ft. Kent, Lewiston, Madawaska, St. Agatha, and Van Buren) had lower mean scores on the MEA than did their monolingual counterparts, the researcher decided to examine the common questions that all pupils must take in mathematics and reading by analyzing the scores of samples from these French and English communities.

Data were analyzed at the individual pupil item level. Scores of identifiably equal pupils, that is pupils from the monolingual English subgroup and from the bilingual/English fluent subgroup with matching identical overall test scores, were compared for item responses. A sample of 1302 monolingual English speaking pupils was required to generate the 336 matching pupil by pupil group by group exact overall test scores with the 336 pupils constituting the entire French bilingual/ English fluent speaking populations of the above referenced communities. Table 7 is provided to indicate that the two subgroups analyzed were of identifiably equal abilities as shown by the identical group

means and standard deviations that could only occur if the pupils from each subgroup were identically matched. Critical to this study is the fact that measured ability (achievement) levels are the same for each of the subgroups analyzed. This fact being established, one can then explore other reasons why differential validity may be occurring between the two subgroups.

TABLE 7

French/English Group Means and Standard Deviations

GROUP	n	Reading Mean	Math Mean	Reading Std Dev	Math Std Dev
FRENCH	336	31.41	24.98	8.06	9.53
ENGLISH	336	31.41	24.98	8.06	9.53

Means and standard deviations for each cultural group.

METHODS OF ANALYSIS

All scores were analyzed by four highly regarded statistical strategies capable of detecting item bias (differential validity), and by a panel of bilingual Franco-American "experts" theoretically capable of doing the same. The following four item analysis strategies, discussed in detail in Chapter Three, were used to determine cross-cultural (French/English) item validity in this study: Scheuneman's modified chi-square procedure (SSX^2); Rudner and Convey's TID-45^0 item difficulty p-value; one-parameter latent trait Rasch model (ICC-1); and the Mantel-Haenszel (M-H) procedure.

VARIABLES

The dependent variable was the MEA test (mathematics and reading common questions), and the independent variable was the language group. The definition of bias was established for the selection process to include items selected by all four statistical strategies. As was described in Chapter Three however, this definition was extended slightly to capture four additional mathematics, and five additional reading items for the expert analysis portion of the study.

REPORT OF DATA ANALYSIS

Each of the item bias methods was used to analyze each mathematics and reading item for the French and English subgroups. Since pupils were matched on the basis of identical overall test scores, item response differences based on group membership can be assumed to indicate item bias, or of the item having or exhibiting qualities of differential validity. In this analysis an item is defined as biased if individuals from different groups, who have the same total score on the test, have different probabilities of responding correctly to that item.

MATHEMATICS STATISTICAL ANALYSIS

Within the bias value parameters established (Table 6), eight mathematics items were selected by the statistical strategies as indicating bias. However, item means-analysis indicated that five items favored the French subgroup while

three items favored the English subgroup. Table 8 below graphically displays the results of the above described analysis. Each of the eight items is followed by an asterisk indicating which combination of statistical strategies selected this item, as well as which group, French or English, was favored by the bias of the item.

TABLE 8

Math Items Selected By Statistical Strategies Indicating Group Favored (G/F)

Item #	SSX^2	$TID\text{-}45^0$	ICC-1	M-H	G/F
3	*	*	*	*	Fr
7		*	*	*	Fr
10	*	*	*	*	Fr
12	*	*	*	*	Eng
19	*	*	*	*	Fr
22	*	*	*	*	Eng
23		*	*	*	Fr
31	*	*	*		Eng

Fr=French Eng=English

DISCUSSION OF CONTENT OF SELECTED MATHEMATICS ITEMS

Table 9 is provided to indicate the process and content categories of the biased eighth grade mathematics items. The process categories assessed pupils' capacity to handle mathematical procedures, concepts, and problem solving skills. The content categories assessed the pupils' ability to manipulate numbers and numeration questions, to handle variables and mathematical relations, and finally to deal with geometry questions. There were no measurement or

problem solving questions identified in the bias analysis either by the statistical procedures or by the experts. Finally, the groups favored by the various items are indicated.

Table 9 shows that of the eight mathematics items identified as having differential validity by the statistical strategies, five (items 3, 7, 10, 19, and 23) favored the French pupil subgroup, and three (items 12, 22, and 31) favored the English pupil subgroup.

TABLE 9

Content Analysis of Statistically Selected Mathematics Items

PROCESS CATEGORIES	**ITEMS FAVORED BY FRENCH**	**ITEMS FAVORED BY ENGLISH**
Procedural	3 and 19	12 and 22
Conceptual	7 and 23	None.
Problem Solving	10	31
CONTENT CATEGORIES		
Numbers and Numerations	3, 7, 10, 19 and 23	31
Variables and Relations	None.	12
Geometry	None.	22
Measurement	None.	None.
Problem Solving	None.	None.

MATHEMATICS ITEM CONTENT ANALYSIS

The following item content analysis examines the French and the English favored items.

FRENCH FAVORED ITEMS

Item number 3 tested for the mathematical procedural process of adding whole numbers with fractions, whereas, item 19 tested for multiplication of whole numbers with fractions. The French pupils in this case handled whole numbers with fractions better than did their English counterparts.

Item number 7 tested for the mathematical concept of rounding large whole numbers to the nearest hundred, whereas, item 23 required ordering decimal numbers from smallest to largest. The French handled these conceptual numeration challenges better than did the English.

Finally, item number 10 tested for word problem solving working with whole numbers and fractions. Again this item indicated that the French handled fractions and whole numbers better than did the English.

ENGLISH FAVORED ITEMS

Item number 12 tested for the mathematical procedure of adding manipulated (multiplied) variables.

Item number 22 tested for the mathematical procedure of figuring the area of a geometric figure. Here again multiplication was involved.

Finally, item number 31 was a word problem involving the manipulation (subtraction) of whole numbers.

COMMON FEATURES

Both groups successfully handled a word problem, this

was the only apparent common feature of the items selected.

NON-COMMON FEATURES

The French subgroup seemed to operate fractions, whole numbers, and multiplication skills more successfully, whereas the English subgroup managed numeric variables, geometry, addition, and subtraction skills more successfully.

CONCLUSIONS

In general, results do little to confirm consistent item bias against the French group. Rather eight items found biased by statistical methods were split five to three in favor of the French group. Neither did content analysis show meaningful differences between the groups.

PANEL OF EXPERT ITEM SELECTION

The panel of experts was asked to examine the items identified by the statistical strategies as indicating differential validity scrambled with an equal number of items not identified as biased. They were asked to select the items they felt were biased and to explain why they believed them to be biased.

Of the eight mathematics items selected as biased by the statistical strategies, only one was selected by any expert as likely to be biased. All remaining items selected as likely to be biased by the experts were selected from among the unbiased items included in the experts' rating task.

Results of the mathematics item analysis done by the panel of experts and compared with items selected by the statistical strategies are provided in Table 10 Each of the eight items selected by the statistical strategies is listed first followed by four additional items identified by the experts. The table indicates if the item was selected by the expert and/or the statistic, which group the item actually favored, and the number of experts and/or statistics which actually selected the item.

It was hoped that the experts would be able to identify most of the biased items identified by the statistical strategies, as expert judgement would be a fairly inexpensive and uncomplicated method to recommend to the State Curriculum Department for screening future items. Unfortunately, the "expert method" in this study had a poor correlation with the "statistical method." With the exception of one item by one judge, the experts were unable to identify any of the biased mathematics items identified in the statistical analysis. Table 10 below illustrates this statistic by expert relationship.

TABLE 10

Statistical Strategies Vs. Experts
Selection Comparisons Mathematics Items

Item #	Expert Selected	Statistic Selected	Group Favored	Number Selecting Exp	Stat
3*	No	Yes	Fr	0	4
7*	No	Yes	Fr	0	3
10*	No	Yes	Fr	0	4
12*	No	Yes	Eng	0	4
19*	No	Yes	Fr	0	4
22*	Yes	Yes	Eng	1	4
23*	No	Yes	Fr	0	3
31*	No	Yes	Eng	0	3
36	Yes	No	Fr	1	0
49	Yes	No	Eng	2	0
35	Yes	No	Eng	1	0
42	Yes	No	Eng	1	0

* = Biased Items Fr=French Eng=English

READING STATISTICAL ANALYSIS

Within the bias value parameters established, nine reading items were selected by the statistical strategies as indicating differential validity. Of these, four items favored the French subgroup while five items favored the English subgroup. This item/group preference is indicated in Table 11. Each of the nine items is followed by an asterisk indicating which combination of statistical strategies selected this item, as well as which group, French or English, was favored by the bias of the item.

TABLE 11					
Reading Items Selected By Statistical Strategies Indicating Group Favored (G/F)					
Item #	SSX^2	$TID\text{-}45^0$	ICC-1	M-H	G/F
10	*	*	*	*	Eng
18		*	*	*	Eng
26		*	*	*	Fr
31		*	*		Eng
33	*	*	*	*	Eng
35	*	*	*	*	Fr
36	*	*	*	*	Fr
37	*	*	*	*	Eng
41		*	*	*	Fr

Fr=French Eng=English

DISCUSSION OF CONTENT OF SELECTED READING ITEMS

Table 12 indicates the process categories and passage types and lengths of the eighth grade reading common items chosen as evidencing differential validity. The process categories covered reading comprehension, handling of reading management strategies, and ability to use reference skills. There were two types of passages used to generate questions in this subtest: literary and practical. Finally, there were both long and short passages. Table 12 also shows which of the nine reading items selected by the statistical strategies favored the French pupil subgroup, and which favored the English pupil subgroup.

TABLE 12		
Content Analysis of Statistically Selected Reading Items		
PROCESS CATEGORIES	**ITEMS FAVORED BY FRENCH**	**ITEMS FAVORED BY ENGLISH**
Comprehension	36	10,18,31,33
Management Strategies	None.	37
Management of Reference Skills	26,35,41	None.
PASSAGE TYPES		
Literary	26	10 and 18
Content	None.	None.
Practical	35,36,41	31,33,37
PASSAGE LENGTH		
Long Passages	35 and 36	10,18,31,33,37
Short Passages	26 and 41	None.

READING ITEM CONTENT ANALYSIS

FRENCH FAVORED ITEMS

Only one reading comprehension item was identified as being differentially valid in favor of the French. This item was based on the long advertisement used in items 31, 33, 37 above. This item required higher order analysis skills. The remaining French-favored items, all involved the management of reference skills in both long and short passages. Item 26 was from a long literary (Frankenstein) passage requiring analysis of the setting of the passage.

Item 35 was from the long advertisement cited above, requiring an analysis of price, shipping and handling charges, and tax in order that a "Total Cost" be established. Finally, item 41 required looking for information in a history book excerpt requiring the pupil to identify the type source the excerpt came from.

ENGLISH FAVORED ITEMS

Item numbers 10, 18, 31, and 33 are all reading comprehension questions. Items 10 and 18 are from long literary passages, whereas items 31 and 33 are from long practical (advertisement) passages. Item 10 deals with understanding metaphor, and 18 deals with higher order analysis. Both items 31 and 33 deal with the comprehension of and decision making based upon advertising hyperbole. Item 37 is based on the same advertising passage as used in items 31 and 33, but involves the utilization of management strategies required in analyzing a graph.

COMMON FEATURES

Both groups dealt successfully with reading comprehension problems from both long and short literary and practical passages.

NON-COMMON FEATURES

The English group had more success with the reading comprehension items (3), than did the French group (1). However, the French group successfully managed reference skill items (3) whereas the English had none. The English

group had success in general reading management strategies (1), whereas the French had none.

CONCLUSIONS

It seems that the French pupils were better skilled at handling reference skills analysis items, whereas the English pupils were better at general reading comprehension items. But in general, results did not confirm consistent item bias against the French group. Rather nine items found biased by the statistical methods were split four to five in favor of the English. Neither did content analysis show meaningful differences between groups.

PANEL OF EXPERTS ITEM SELECTION

The panel of experts was asked to examine the items identified by the statistical methods as indicating differential validity scrambled with an equal number of items not identified as biased. They were asked to select the items they felt were biased and to explain why they believed them to be biased.

Of the five reading items selected by members of the Panel of Experts, all five were also selected by the statistical strategies. In addition, the experts selected five more items from among the unbiased items included in the experts' rating task.

Results of the reading item analysis done by the panel of experts and compared with items selected by the statistical strategies is provided in Table 13. Each of the

nine items selected by the statistics is listed first followed by five additional items identified by the experts. The table indicates if the item was selected by the expert and/or the statistic, which group the item actually favored, and the number of experts and/or statistics which actually selected the item.

As was stated in the mathematics analysis section, it was hoped that the experts and the statistical strategies would have a fair degree of correlation in selecting potentially biased items. The "expert method" did operate more effectively in the reading section, where the experts were correct in their identification about 56% of the time. However, if the selection of "non-biased" items is included, the percentage of correct choices by the experts drops to only about 36%

TABLE 13					
Statistical Strategies Vs. Experts Selection Comparisons Reading Items					
Item #	Expert Selected	Statistic Selected	Group Favored	Number Selecting Exp	Stat
10*	Yes	Yes	Eng	1	4
18*	No	Yes	Eng	0	3
26*	No	Yes	Fr	0	3
31*	Yes	Yes	Eng	1	2
33*	Yes	Yes	Eng	1	4
35*	Yes	Yes	Fr	3	4
36*	Yes	Yes	Fr	1	4
37*	No	Yes	Eng	0	4
41*	No	Yes	Fr	0	3
24	Yes	No	Equal	1	0
34	Yes	No	Fr	1	0
11	Yes	No	Fr	1	0
42	Yes	No	Fr	1	0
13	Yes	No	Fr	1	0

* = Biased items Fr=French Eng=English

NULL HYPOTHESES

The null hypotheses for the analysis stated that there will be no difference in the probabilities of item level success on the 1989, common MEA mathematics and reading items between the two subgroups: the French bilingual/ English fluent pupil group, and the monolingual English pupil group.

The null hypotheses were not rejected for either the reading or the mathematics items. Although in fact eight

(8) mathematics items, and nine (9) reading items were identified by the statistical strategies as indicating differential validity, five of the eight mathematics items actually favored the French subgroup, while three favored the English subgroup, and five of the nine reading items favored the English subgroup while four favored the French subgroup. Such results render unidirectional bias concerns virtually a moot point in this study. Whereas differential validity may well be a factor in all of these items, the net effect is to favor neither group materially, the conclusions cannot be drawn on the basis of this study that the MEA as a whole is biased in favor of the English subgroup.

CONCLUDING DISCUSSION OF OVERALL FINDINGS

Whereas, at the outset of this study, it was expected that items in the MEA might be biased in favor of the French subgroup, statistical analysis indicated the MEA actually marginally favored the French subgroup. It is clear, however, that item level statistical analysis of tests like the MEA is essential to investigate test item validity as well as to maintain the high levels of public confidence envisioned by the legislature.

For purposes of subgroup identification convenience, this study was of the French/English subgroups. These two language groups have many common evolutionary/developmental (etymological) features. Additionally, it must not be overlooked that the lower mean scores of the so-called

bilingual/English fluent pupils of the state that prompted this study in the first place, surely reflect a significant factor of mis-identification of just what constitutes a fluent/bilingual pupil as was indicated in the administrator survey found in Appendix B. That is, pupils who are truly bilingual/English, and who have been determined to be so by techni lly defensible selection processes, are almost certainly not going to generate lower mean scores on a state-normed test like the MEA. In fact as the research has indicated, they might in fact generate higher mean scores than their monolingual English counterparts. These points, combined with the fact that pupils' item response comparisons were paired in terms of overall test scores, undoubtedly played a role in the discovery that item differential functioning was at play in the MEA, and hence, overall test bias could not be established in favor of one group over the other.

Further, from a technical standpoint, the statistical analysis strategies used in this analysis indicated a high degree of compatibility. H. Johnson Nenty's study demonstrated a similar high level of statistical strategy correlation. The point here is that it is probably not necessary to analyze pre-tested items in the MEA (matrix-sampled items) with more than one statistical method. The Mantel-Haenszel Procedure is probably the easiest and least expensive method to use. Such a finding will make item

analysis by the Curriculum Division of the Department of Education less burdensome, and certainly less costly.

Finally, it is clear that there was little overlap between the statistical methods used and the conclusions drawn by the experts on the panel of experts. The hoped for correlation between these two methods of analysis for determining differential validity did not occur at least with the methodology devised for this study.

CHAPTER FIVE

Summary, Conclusions, And Recommendations

SUMMARY OF RESEARCH FINDINGS:

An analysis of the 1989, eighth grade common mathematics and reading items on the Maine Educational Assessment Test indicated that eight out of fifty mathematics items (16%) and nine out of fifty reading items (18%) from the common questions asked of all eighth grade pupils in these two subject areas were selected by four widely accepted statistical strategies routinely used for identifying item bias in research applications. These statistical strategies: Scheuneman's modified chi-square procedure (SSX^2), Rudner and Convey's TID-45^0 item difficulty p-value, the one-parameter latent trait Rasch model (ICC-1), and the Mantel-Haenszel procedure showed a high degree of compatibility in item selections reported in this study. (The chi-square values, Z-scores, and item p-value differences generated by these statistics are reported in Appendix C of the study).

Additionally six items 35% were selected from the same subgroup of seventeen items by French bilingual/English fluent experts as also indicating differential validity from their own personal frames of reference. In short, this study concludes that a significant percentage (± 17%) of the common mathematics and reading questions on the 1989, eighth grade MEA test showed indicators of item bias.

If by extension one were to look at the whole data bank of questions put before Maine's children in 1989...889 items across all subject area disciplines, (216 mathematics items and 224 reading items,)...and were to apply the same percentages suggested in this study as indicating possible differential validity (item bias), the number of items so distinguished is quite remarkable and is shown in Table 14.

TABLE 14

Hypothetical Extension To All 889 MEA Test Items of % of Biased Items Found In The 100 Common Reading and Mathematics Items

Data bank questions:	% biased	# biased	Hypothetical # biased by extension
889 items all subjects	17 % hypothetical [70]	not actually analyzed	151
224 reading (50 common)	18% (common)	9	40
216 math (50 common)	16% (common)	8	35

RESEARCHER'S CONCLUSIONS

There can be little doubt that in general the MEA has had a salutary impact on most of Maine's schools. This fact is especially true concerning the writing section of the test. This research is clear, however. Bilingual/English fluent minority children could be impacted by biased items in the MEA, although this research finds that such

[70] The average of 16% and 18% the actual percentages found in the reading and mathematics analysis.

potential bias is moot with respect to the French bilingual/English pupils studied here. Nevertheless, when decisions are to be made on the basis of tests like the MEA, all efforts must be made to insure that all bilingual/ English fluent children are treated the same as the monolingual English children also taking the test, regardless of language of birth , or culture of origin. Furthermore, if tests cannot be created to treat such bilingual/English fluent populations equitably, then such pupils should be tested in separate settings with instruments appropriate to their needs, or they should be exempted from the consequences and the publicity of such testing altogether. To ignore such research findings would be to deny both the multicultural nature of our society as well as to deny the obligation psychometricians inexorably have to serve equitably all cross-cultural populations in Maine's schools.

RECOMMENDATIONS

Assuming that political or legal pressures under the public's right-to-know provisions keep constraints on the Education Department's ability to alter MEA reporting procedures, item screening strategies must be explored to insure that differential validity is minimized or eliminated for language minority pupils.

Notwithstanding the conclusion that the items analyzed in this study do not constitute significant bias in favor of

one group over the other, the fact remains, that by definition, items selected by the statistical strategies as indicating differential validity fit the profile of item bias established in this study's methodology, and supported by its literature review. Given the fact that there are dozens of language groups in Maine, and thousands of pupils from cultures other than European/British/American English-speaking backgrounds, cross-cultural validity issues in the MEA remain to be settled before each presentation of the test. For almost certainly, bilingual/English fluency issues concerned with cross-cultural item bias will continue to present problems for bilingual pupils. This researcher concludes that cross-cultural linguistic item bias will be a continuing factor requiring item analysis, further item screening, and some item replacement at least for the major language groups of the state: French, Asian, German, Spanish, American Indian, and others as determined from time-to-time. What to do about testing equity for other more marginal language groups is certainly a topic that Maine's testing experts, and the consultants they hire, will need to consider if justice and fairness are to prevail in Maine's MEA testing program.

This study explored three areas that should be explored by Maine's Department of Education to better serve language minority pupils with respect to the MEA in future testing cycles:

First, pupil language competency.

No child should be tested with the MEA instrument until it is absolutely clear that s/he has been given appropriate language assessment prior to taking the MEA to clearly establish the requisite English language skills to compete equitably for success in this test.

Second, statistical item analysis.

Item analysis with a statistical strategy like the Mantel-Haenszel procedure should be routinely run on the matrix-sampled items being considered for the common mathematics and reading items. Outlying items which indicate cross-cultural differential validity should be edited or rejected as a matter of routine procedure.

Finally, expert item analysis.

The Department of Education must recruit and train assistants, from the major language groups outlined above, to work with the test developers in screening out test items with potential cross-cultural differential validity (item bias) problems.

Clearly the survey results provided in Appendix B indicate that the guidelines outlined in "appropriate educational practices for limited English proficient students" (Appendix A) are generally applied on the subjective end of a process continuum, rather than on the more formal objective end. In fact, it was noted in Chapter Two that fully 63% of Maine's school administrators

responding to the Appendix B survey used "conversational English proficiency through subjective observation" to determine if a pupil was a bilingual/English fluent speaker and hence competent to take the MEA. Research cited in this study has substantiated the problematical nature of such casual conversational approaches to determining language competency for technical educational purposes.

This researcher believes that the Division of Curriculum (now responsible for administering the MEA in the State) should establish rigorous procedures mandated of all school administrators for such language competency determinations. Furthermore, school administrators making such determinations based on such new " rigorous procedures," should be required to sign sworn statements, declaring penalties for perjury, in order to give this critical component of the MEA program the status needed to protect the rights of the language minority pupils. The casual, slip-shod methods currently employed by well over half of the administrators responding to the referenced survey are almost certainly having a deleterious effect on the pupils concerned, and has thrown into question this aspect of the MEA program for citizens of Maine who are informed and concerned about such matters.

This study utilized experts fluent in the language of the pupils' birth as well as in English to vet items for possible bias. This approach proved to be far from exact,

however, producing at best only about a 50% agreement correlation with the statistical strategies. It is perhaps better than no cross-cultural item bias analysis at all, though, which is currently the case with this MEA test. However, as recommended above, language-competent assistants from the major language groups of the state, must be well and technically trained to spot item bias issues as they work with the test developers attempting to deal with items flagged by the recommended statistical analysis procedure. Not to provide such technical training would be to seriously reduce the efficacy and reliability of such an item-bias analysis process.

In the current study solo experts were asked to read and analyze test items, and did not communicate with other members of the panel of experts on the substance of the items. A number of these experts did communicate about the process, however, and felt that a team approach to such an expert analysis process might prove more efficacious. Such an approach is apparently often used in test standardization procedures.

As recommended above, expert training in the identification of items potentially capable of producing differential validity should be provided those people vetting test items for possible bias in all the major language groups covered by tests like the MEA. The fact that French bilingual/English fluent pupils in Maine were

apparently not impacted by the questions analyzed in this study is remarkable. In fact some items in the test actually favored the French subgroup. However, other studies will have to look at other non-English cultural group performances to make certain that pupils in other than French cross-cultural groups are not being disenfranchised of some social or educational benefits based on the MEA. In fact it cannot be certain that even the French population would not be negatively impacted in future testing cycles that had not been cross-culturally standardized as recommended in this section.

It is true that MEA test items are pre-tested through the matrix sampling process, and that item curves are analyzed, before items are included in their one-shot-performance in the common question bank. Certainly this process eliminates some outlying items, and insures a higher quality item in general in the common question bank. However, such a process gives little comfort to multiculturalists, when it is learned that no cross-cultural validation, whatsoever, takes place in the development of the MEA. Given the complex multicultural nature of our society today, this fact is mind-boggling to say the least.

With one hundred language groups currently extant in Maine, and with over 6000 pupils functioning on a continuum from Limited English Proficient (LEP) to bilingual/English fluent, test developers and State curriculum experts must

recognize that they have a responsibility to equitably serve the multicultural pupil population in the State's objective-referenced test, as well as to serve the monolingual English pupil population.

AFTERWORD

A headline in the March 20, 1991, "Education Week" newspaper stated: "[1990] Census Confirms Remarkable Shifts In Ethnic Makeup...Cultural Diversity Called Challenge for Educators." [71] Here are a few of the main points of this article:

1. … the nation underwent a remarkable transformation in the last decade…
   a. … Asian-American population doubled…
   b. … Hispanic population up by 50%…
2. … a nation increasingly culturally diverse…
   a. 199.7 million whites, an increase of 6%
   b. nearly 30 million Blacks up by 13.2%
   c. nearly 2 million Native Americans and Eskimo-and Aleut-Americans up by 37.9%
   d. 7.3 million Asian- and Pacific Islander-Americans up by 107.8%
   e. 22.4 million Hispanics up by 53% [72]

Such data indicate that there will be, and most

---

[71] Liz Schevtchuk Armstrong, "Education Week," (Washington, D.C.: Editorial Projects in Education, Volume X, Number 26, March 20, 1991), 1.

[72] Ibid, 16.

assuredly ought to be, many pressing issues in American educational circles around the cross-cultural concerns of bilingual/multicultural education and the rights of children from language mincrity populations. Surely non-biased cross-cultural assessment ought to be a topic uppermost in the minds of psychometricians as they strive to better serve their increasingly multiculturally diverse clientele.

This researcher examined bias in testing as an impact area within the paradigms of cultural pluralism and multicultural/bilingual education as such bias might impact tests like the MEA. Given the compelling demographics outlined above in the census synopsis…the culturally pluralistic framework of our society as it were…one would hope for high levels of cultural and linguistic sensitivity among the experts working to develop tests appropriate for all the pupils of the State.

Since upwards of forty states are now mandating similar tests, and indeed, since the federal government is considering a similar national educational accountability test, it seems appropriate to explore the question of how the language minority status of many thousands of students is impacted, performance-wise, on school accountability instruments like the MEA which have achieved such a high profile in the United States in such a relatively short historical time period; that is, since the publication of "A Nation At Risk."

Moreover, there is significant consideration raised by Elliot Eisner, that has a collateral part to play in this study. That is the confusion between evaluation and measurement. Tests, like the MEA, grew out of a political need on the part of members of the state's legislature to hold schools accountable for declines in standardized testing scores over the past twenty years; that is to say to evaluate the performance of the State's schools. But, bona fide evaluation processes are time consuming, expensive, and often inconclusive... prerequisites politicians are ill-equipped to deal with. Measurement, conversely, is much quicker and less complicated--qualities appealing to politicians. The MEA is a compromise instrument which attempts to bridge the chasm between evaluation and measurement.

Eisner sees the dominance of the scientific view reflected in educational measurement as having a "deleterious" effect on the curricula offered in America's schools. He points out that we can evaluate without measuring, and that we can measure without evaluating; but there has come to be a dominant thinking on the part of members of the state legislature, that "one must measure in order to evaluate."

> For the curriculum of the school this means that evaluation practices, determined largely with respect to what can be measured, influence to a very large degree the kinds of programs that will be offered to the young. Educational practices based on a scientific model too often become not a tool for

> improving the quality of teaching and learning but rather an impediment to such ends. Students and teachers alike gear up to take tests, even though none of them believes those tests to be intrinsically important or that the tests really assess much of what has been learned and taught in schools.[73]

Maine's educational programs are both served and at the same time perhaps ill-served thrcugh the use of the Maine Educational Assessment testing program…a program put in place by a legislature which may have been confused by the very different essences of the words "measurement" and "evaluation."

If tests like the MEA are to be mandated, then the legislature must be made aware of the fact that educational experts responsible for carrying out such mandates have a moral obligatic:n to be concerned about "…the broader concerns of social justice and the appropriateness of test use for groups affected by testing."

With the self esteem of so many culturally diverse children at stake, along with the cultural and linguistic integrity of so many non-English Maine families, such a study as this may contribute positively to the body of knowledge indicating that careful thought and planning need to be a part of Maine's educational "evaluation" movement with respect to cross-cultural assessment in specific. For the political demands for accountability need not sacrifice

---

[73] Elliot W. Eisner, The Educational Imagination (New York: Macmillan Publishing Company, 1985), 14-15.

the overriding requirements for sound pedagogy and equitable assessment.

Appendix A

Appropriate Educational Practices For Limited English Proficient Children.

John R. McKernan, Jr.
Governor

Eve M. Bither
Commissioner

DEPARTMENT OF
EDUCATIONAL AND CULTURAL SERVICES
Telephone (207) 289-5800

ADMINISTRATIVE LETTER NO. 28
POLICY CODE: CGC

TO: SUPERINTENDENTS

FROM: Eve M. Bither, Commissioner

RE: APPROPRIATE EDUCATIONAL PRACTICES FOR LIMITED ENGLISH PROFICIENT STUDENTS

DATE: MARCH 17, 1988

In the past few years, Maine school districts have enrolled an increasing number of students who speak little or no English. Many of these students are refugee children from Southeast Asia, Afghanistan, Iran, and Poland. Many school districts have found it difficult to identify and implement appropriate and effective educational programs for this low incidence, limited English proficient student population.

Obviously, these new American students are entitled to equal access to the American educational system. Educational services should provide English language communication features, learning skills, and subject content to prepare the limited English proficient student for academic achievement in all English instruction in the classroom. Inadequate language and skills development instruction as well as premature exiting from specially designed programs result in academic failure for those students not quite ready for English-only content work in a majority of cases.

The attached document is intended to provide assistance to Maine school districts in designing and providing effective instructional programs to low incidence, limited English proficient students. First, it defines the legal obligations of the school district to this special population. The remainder of the document offers guidelines for fulfilling these obligations with appropriate language and learning skills development programs.

Inquires may be directed to Dr. Barney Berube or Mr. Val Hart through the office of federal projects for minority languages at 207 289-5980.

APPROPRIATE EDUCATIONAL PROGRAMS FOR LOW INCIDENCE, LIMITED ENGLISH PROFICIENT STUDENTS

INTRODUCTION

The United States Congress addressed the subject of discrimination against linguistic minority students in the Equal Education Opportunity Act of 1974, 20 U.S.C. 1702.

> No state shall deny equal educational opportunity to an individual on account of his or her race, color, sex or national origin, by ...(f) the failure by an educational agency to overcome language barriers that impede equal participation by its students in its instructional programs.

Congress acted to ensure that all public schools should comply with this act, not just those receiving federal funds. This statute recognizes the state's role in assuring equal educational opportunity for national origin minority students. The statute also stresses that the failure of an educational agency to rectify appropriately a limited English proficient student's English competencies is a denial of equal educational opportunity.

To be consistent with current educational nomenclature, this document will refer to such students as Limited English Proficient (LEP) students. This terminology recognizes that some LEP students may have learned a degree of English understanding and speaking abilities but remain limited in reading and writing English. Also, the LEP nomenclature is mindful that limited English proficient students need more than just aural-oral abilities to achieve in English language standard curriculum instruction.

The guidelines which follow are consistent with federal decisions of the United States Supreme Court and significant lower court decisions, memoranda from the United States Department of Education and the Office of Civil Rights, Washington, D.C. as well as a review of the theoretical and empirical research literature in the fields of second language acquisition, bilingual, and English as a second language education.

LEGAL REQUIREMENTS

A. School systems must identify all students whose primary language is other than English, who have or may have difficulty performing ordinary classwork in English, and who can not learn or achieve on parity with their English dominant peers. Such LEP students must be placed in a specifically designed language support program. (Lau guidelines)

B. Any specially designed support or instructional program shall be consistent with all federal acts and mandates, related federal regulations and court cases as well as Maine State acts, mandates and policies, which relate to the education of limited English and National Origin Minority students.

C. This instructional program should be based on second language acquisition pedagogy and sound educational practices for meeting the individual needs of LEP students. The burden of proof is upon the district that the instructional program designed for a LEP student has clearly developed English language skills of comprehension, speaking, reading and writing necessary for learning and achieving in English-only instruction at a level substantially equivalent to pupil whose primary language is English (Castaneda v. Pickard, 648 F2nd 989-5th Circuit-1981)

D. A school system shall document all procedures and activities especially designed for Limited English Proficient students enrolled in its schools (Chapter 71-A, Section 5.). A formal policy statement and description of procedures for identifying, language-level classifying, instructional placement, reclassifying and/or mainstreaming LEP students should be available to appropriate staff members and parents/guardians.

E. The identification, diagnosis of language proficiency, and instructional placement of Limited English Proficient students must be conducted by a certified and qualified bilingual or English as a second language teacher, or by a certified counselor trained in the language and pedagogical procedures of bilingual education. (MRSA 20 ch.118)

F. School systems which provide English language development programs to LEP students should reclassify students from limited English proficient (LEP) to fluent English proficient (FEP) by specific multi-criteria reclassification procedures-(Rios v. Read, 73

F. R.D., 595 (E.D. N.Y.-1977) Cintron v. Brentwood, E.D.N.Y. #77-C-1370.

PROCEDURES FOR IDENTIFICATION AND LANGUAGE PROFICIENCY CLASSIFICATION OF LIMITED ENGLISH PROFICIENT STUDENTS

A. To facilitate the identification and acceptable program placement of LEP students, the district should identify a suitable person or Language Assessment Committee to coordinate and oversee the educational program of LEP students enrolled in the school system. The person(s) will

1. conduct identification and language classification assessment activities;
2. assure appropriate program and instructional placement of student(s) classified as English proficient;
3. meet periodically with relevant bilingual, English as a second language, and standard curriculum staff to determine if student(s) is ready for partial or full-time mainstreaming;
4. develop and implement appropriate procedures for language proficiency reclassification of bilingual students;
5. monitor the follow-up activities for partial and fully mainstreamed students;
6. make recommendations for instructional or other services for partially and fully mainstreamed students;
7. develop a process for informing relevant bilingual, English as a second language or standard curriculum staff of LEP student progress;
8. establish a record keeping system for recording assessment results, instructional placement, reclassification procedures, and follow-up monitoring activities.

B. Identifying Newly Registering LEP Students

1. Identify primary/home language of the student(s) with "Lau" Home Language Survey. The Department conducts this process annually.
2. Interview the students and/or parent(s) or legal guardians(s) in the primary/home language to determine garde level and academic experiences, native language learning experiences the student has had.
3. Review by Language Assessment Committee all available educational documents or credentials. Relate this data to available English language descriptions of foreign educational procedures.
4. Determine English language proficiency using acceptable procedures and instruments, administered by qualified certified bilingual or

English as a second language instructor.

a. For kindergarten and primary level students: access student's listening and speaking competency, and English language readiness skills
b. For grades 3 through 12: access student's aural-oral skills, and reading and writing abilities.

C. Identifying Currently Enrolled LEP students

1. Implement a standard curriculum classroom survey to identify students:
   a. who have a primary/home language other than English (home language surveys);
   b. who are not functioning on grade level;
   c. whose lack of academic achievement is due to limited English language proficiency.

2. Determine English language proficiency using acceptable procedures and instruments, administered by a qualified and certified bilingual or English as a second language instructor. Assessment instruments and procedures should evaluate English listening/understanding, speaking, reading and writing abilities. Student achievement should be comparable to English-language of the same age and grade level.

INSTRUCTIONAL PLACEMENT FOR STUDENTS IDENTIFIED AS LIMITED ENGLISH PROFICIENT

A. Based upon the language and educational data collected by the

1. Home language survey or standard classroom survey,
2. Native language interview,
3. Review of available documents or credentials, and,
4. English language proficiency assessment activities

the limited English proficient student shall be placed in one of the following instructional programs.

B. A limited English proficient student will be placed in

1. the appropriate grade or level or instruction in a full-time transitional bilingual education program

or

2. a native-language supported English language development program

or

3. a structured English as a second language development program.

APPROPRIATE PROCEDURES FOR RECLASSIFICATION AND TRANSFER OF LIMITED ENGLISH PROFICIENT STUDENTS

A. Districts must establish criteria for reclassifying a student's language proficiency before transferring the student to another instructional program.

1. Determine criteria which a LEP student would need to meet if she/she is to be reclassified as fluent English proficient (FEP). Criteria should be determined for fluent proficiency in English language skills of comprehension, speaking, reading and writing.

2. Identify assessment instruments and activities which are linguistically and culturally relevant for testing English proficiency domains.

3. Districts are urged to supplement language assessment activities with additional data on student achievement and other relevant factors for ensuring LEP students are prepared to receive instruction only in English.

B. A district should adopt specific criteria re-classification procedures consisting of

1. Teacher evaluation: relevant instructional staff would evaluate a student's general language proficiency by observing the student's oral performance in several formal and informal settings, and completing an observation-matrix profile.

2. Objective evaluation of a student's mastery of English as a second language skill objectives: an ESL instructor would administer a criterion-referenced test for evaluating mastery of ESL features and skills.

3. Objective assessment of oral language proficiency: in an appropriate and valid English proficiency instrument(s), the student must at least meet the publisher's specified English speaking fluency score.

4. Objective assessment of English language arts abilities: including the four skills area of listening, speaking, reading and writing, the student must demonstrate achievement on parity with their English peers at the same age and grade level.

C. After implementing reclassification procedures, the language assessment committee would determine a new part or full time instructional placement for each student. A re-placement option would be placement in a standard curriculum classroom or program commensurate with the student's chronological age and grade level with daily or frequent English language support services - e.g., reading and writing instruction, content area tutoring, or advanced communication competencies instruction.

D. Within two weeks of a new instructional placement, a district should follow-up and review a reclassified student's academic achievement and psycho-social adjustment. The aim of this follow-up review would be to ascertain if the reclassified student is able to academically compete with English language peers in all-English instruction. Another aim is to determine if the student is adjusting socially and psychologically to the new instructional placement.

E. The school district should establish procedures for periodically monitoring the newly classified fluent-English proficient student for at least three (3) years after reclassification and mainstreaming. The objective of this monitoring process is to objectively prove that the non-English language background student is adjusting to the new instructional setting and functioning substantially equivalent to non minority students in English-only instruction. Thus, the school district must insure that they continue to address linguistic, academic and psychosocial needs of non-English language background students.

F. The district or language assessment committee will notify the parent(s) or guardian(s) of the student of the re-classification and new instructional placement.

1. Written notification should be in the primary language of the parent.
2. Notification to the parents should include information of their legal right to challenge the reclassification of their child.

RECOMMENDED TEACHER SKILLS

Regardless of the organization of an English program for LEP students, teachers will find the following competencies important to successfully providing appropriate services. English language development (ELD) teachers should have:

1. insight into the cultural and linguistic backgrounds as well as the academic experiences of their pupils;
2. a special knowledge of the sounds, syntax, semantics, and prosody in their own language to help their students learn these features as well as conflicts between the two language features;
3. experience in the methodology for successful second-language instruction;
4. contact with human or agency resources available to teachers of non-English-speaking students;
5. knowledge of the current issues and research in second-language pedagogy-psycholinguistics, bilingual education, and applied linguistics;
6. awareness of the dynamics of individualized instruction and how to use them in ESL instruction;
7. skills for adapting and developing relevant instructional materials for ESL instruction;
8. familiarity with ways to provide a learning atmosphere that promotes not only second-language skills but also enhances the sense of self in second-language students.

All these skills are necessary because ESL teachers often must provide instruction to students with a wide range of abilities, experiences and levels in both their native and second language; immigrant students sometime register continuously through the school year; ESL teachers often are liaisons between the non-English-speaking students and their English-language teachers.

Recent studies substantiate the existence of a positive relationship between English language proficiency and academic achievement measured in English. These studies and others establish the necessity of a level of quality and quantity of English competency plus academic skill necessary for achievement in English-only instruction.

LANGUAGE ASSESSMENT CONSIDERATIONS

A low level of English proficiency is not the only cause of under achievement among LEP and NELB students. Other variables affect a student's success: lack of equal access to special and tutorial services, institutional barriers, school social climate, perceived academic status, lowered socio-economic goals and previous academic experiences. Consequently, districts should consider the long-term affects of the instructional program provided LEP students, and the extent to which academic achievement is sustained after reclassification to FES takes place.

Using English language achievement instruments of proven validity and reliability, a district needs to assess a transitional LEP student's skills in comprehending-listening, speaking, reading, and writing domains of English language arts. The aim of this testing is to determine if the LEP student is able to learn with the English language in all-English instruction of the content areas in the standard classroom.

(1) If a norm-referenced test is used, the student's performance is compared to district non-minority norms. If district norms are low, the LEP student's performance is compared with National non-LEP norms.

(2) If a standardized criterion-referenced test is used, tests items should represent language arts skills in which English fluent pupils of the same grade and age are expected to be proficient.

(3) The score or percentile which the LEP student is required to meet must represent comparable ability by his/her English dominant peers. That is the cut-off scores chosen are relatively equivalent to the performance of English fluent students of the same age and grade.

(4) Scoring options might be:

a. cut-off scores not less than 36th percentile of the norming group scores (students scoring at or above the 36th percentile would be reclassified as "fluent English proficient" if teacher evaluation and oral-proficiency criteria are also met.)

or

b. cut-off scores in the range of the 31st to

the 35th percentiles of the norming group chosen. (Students scoring within this range would be exited only with parent approval and if the committee judges the student's English language ability sufficient enough for successful learning in English-only instruction but with appropriate language support services as needed.)

or

c. cut-off scores for standardized criterion-references tests are relatively equivalent to the average performance of English dominant students of the same age and grade. (Students scoring at this level would be reclassified fluent English proficient if teacher-evaluation and oral-proficient criteria are also met.)

Local educational agencies are financially pressured to mainstream LEP students as soon as possible. However, the long-range effects of premature reclassification and mainstreaming often result in poorer achievement by LEP students in all-English instruction. Districts then need to spend more monies on remedial services.

REFERENCES

Cummins, James, "Linguistic Independence and the Educational Development of Bilingual Children, "Review of Educational Research, Spring, 1979, 49(2), 222-251.

_____, "The Role of Primary Language Development in Promoting Educational Success for Language Minority Students," Schooling and Language Minority Students: A Theoretical Framework, Evaluation, Dissemination and Assessment Center, California State University, Los Angeles, CA 1981.

DeAvila, Edward A., Cervantes, R.A. and Duncan, S., "Bilingual Program Exit Criteria," CABE Research Journal, 1978, 1 (2), 23-29.

Duncan, Susan E. and DeAvila, E.A., Bilingualism and Cognition: Some Recent Findings, "NABE Journal, Fall, 1979, IV (1), 15- 50.

Krashen, Stephen D., Second Language Acquisition and Second Language Learning, London: Pergamon Press, 1981.

Appendix B

Survey Of Maine School Administrator.

E X P L A N A T I O N

On the Maine Educational Assessment, a subgroup reporting question relates to English language fluency. The examiner's manual offered the following description to help administrators respond to this subgroup part.

Which Category of English language fluency best describes the student? (Mark only one)

_____ A student is monolingual English if there is no evidence that a language other than English is used in the student's home environment. The Department of Education Lau home language surveys were used to determine if students were bilingual or monolingual.

_____ A student is **bilingual/English fluent** if there **is** evidence that a language other than English is used in the student's home environment **and** the student is proficient in **all** English communicative skills areas of reading, writing, speaking and listening commensurate with his/her monolingual English peers. The Department of Education Lau home language surveys were used to determine if students were bilingual or monolingual.

_____ A student is **bilingual/limited English proficient** if there is evidence that a language other than English is used in the student's home environment **and** the student has **limited English proficiency in one or more** of the English communicative skills areas of reading, writing, speaking, or listening.

Guidelines for assessing limited English proficiency were published in the Commissioner's Administrative Letter #28 (March 17, 1988).

Please RETURN YOUR RESPONSE TO THE FOLLOWING QUESTION AS SOON AS POSSIBLE TO: **Dr. Barney Berube, Maine Department of Education, Division of Curriculum, State House Station No. 23, Augusta, ME 04333.**

In selecting the MEA subgroup response "bilingual/English fluent," I used the following determinant(s):

_____ The results of language assessment instrument (e.g., Language Assessment Scales, Language Assessment Battery, Individual Proficiency Test,

Macaulitis...)

____ The results of an intelligence test

____ The results of a standardized achievement test

____ Academic grades

____ Conversational English proficiency through subjective observation

____ Writing competency from in-class observation

____ Reading competency from in-class observation

____ The results of a language assessment committee's documentation

____ Student self-identification as "bilingual/English fluent"

____ Home language survey indication of bilingualism

____ The subjective judgment of another staff person(s)

____ The subjective judgment of the parent(s)

____ Other:_____

____ I do not know how the determination was made.

MEA SURVEY RESULTS*
December 1990

In selecting the MEA subgroup response "bilingual/English fluent," the following determinants were used:

	N	%
The results of a language assessment instrument (e.g., Language Assessment Scales, Language Assessment Battery, Individual Proficiency Test Macaulities...)	15	24
The results of an intelligence test	2	3
The results of a standardized achievement test	16	25
Academic grades	14	22
Conversational English proficiency through subjective observation	40	63
Writing competency from in-class observation	34	54
Reading competency from in-class observation	32	51
The results of a language assessment committee's documentation	5	8
Student self-identification as "bilingual/ English fluent"	19	30
Home language survey indication of bilingualism	32	51
The subjective judgment of another staff person(s)	34	54
The subjective judgment of the parent(s)	13	21

*Based on 63 returns out of 91 distributed (69% returned)

Other:

(a) In prior years the ESL teacher determined the English language fluency of the students who were served by her.

(b) ESL testing

(c) Knowledge of parents, ESL adult programs; some of the children identified are not in my opinion, bilingual, or limited English.

(d) No students in our grade four multilingual class were identified as "bilingual/English fluent"; the above pertains to any students so identified fully mainstreamed in all classes.

(e) This subgroup response was not checked at my school (while it may have been in my district).

(f) This choice varies according to the student, whether or not the student was receiving services with the ESL tutor or had graduated from the program as well as from observations.

(g) Archer Israel - ESL teacher

I do not know how the determination was made. Total 0

Forms returned with no response 1

Appendix C

Mathematics And Reading Statistics By Strategy

TABLE 1 A

MATHEMATICS (SUBTEST ONE)

The French bilingual/English fluent pupils were examined for item responses compared with the monolingual English pupils on the same items.

Items in SUBTEST ONE indicating possible differential validity as indicated by Scheuneman's modified chi-square procedure: SSX^2

ITEM NUMBER	CHISQ
3 *	6.73513
10 *	7.39737
12 *	4.97304
14 -	7.58636
19 *	5.96622
22 *	7.80167
24 -	5.10044
31 #	4,53252
49 -	7.09026

* = selected items common to all four statistical strategies.

- = selected by SSX^2 only

= selected by a total of three strategies absent M-H

TABLE 2 A

MATHEMATICS (SUBTEST ONE)

The French bilingual/English fluent pupils were examined for item responses compared with the monolingual English pupils on the same items.

Items in SUBTEST ONE indicating possible differential validity as indicated by Rudner & Convey p-values analysis: TID-45^0. This subtest contains 50 items.

ITEM NUMBER	P-VALUE DIFFERENCE
3 *	0.31988
7 #	0.30044
10 *	0.42390
12 *	0.28454
19 *	0.24412
22 *	0.27962
23 #	0.25788
31 #	0.37461
42 -	0.25700
43 -	0.26746

* = selected items common to all four statistical strategies.

- = selected by TID-45^0 only

= selected by a total of three strategies absent SSX^2

TABLE 3 A

MATHEMATICS (SUBTEST ONE)

The French bilingual/English fluent pupils were examined for item responses compared with the monolingual English pupils on the same items.

Items in SUBTEST ONE indicating possible differential validity as indicated by the Rasch one parameter latent trait model: ICC-1. This subtest contains 50 items.

ITEM NUMBER	Z SCORE
3 *	1.53569
7 #	1.02584
10 *	1.72088
12 *	-1.37917
19 *	1.74131
22 *	-1.08134
23 #	1.33989
31 #	-1.64855

* = selected items common to all four statistical strategies.

= selected by a total of three strategies: absent M-H for item 31, and absent SSX^2 for items 7 and 23

TABLE 4 A

MATHEMATICS (SUBTEST ONE)

The French bilingual/English fluent pupils were examined for item responses compared with the monolingual English pupils on the same items.

Items in SUBTEST ONE indicating possible differential validity as indicated by the Mantel-Haenszel Procedure: M-H. This subtest contains 50 items.

ITEM NUMBER	CHISQ
3 *	0.025
7 #	0.044
10 *	0.005
12 *	0.044
19 *	0.035
22 *	0.073
23 #	0.062

* = selected items common to all four statistical strategies.

= selected by a total of three strategies absent SSX^2

TABLE 5 A

READING (SUBTEST TWO)

The French bilingual/English fluent pupils were examined for item responses compared with the monolingual English pupils on the same items.

Items in SUBTEST ONE indicating possible differential validity as indicated by Scheuneman's modified chi-square procedure: SSX^2.

ITEM NUMBER	CHISQ
10 *	4.70826
33 *	5.84153
35 *	3.32207
36 *	6.17727
31 ~	3.40376
37 *	3.78645

* = selected items common to all four statistical strategies.

~ = selected by a total of two strategies absent ICC-1, and M-H.

TABLE 6 A

READING (SUBTEST TWO)

The French bilingual/English fluent pupils were examined for item responses compared with the monolingual English pupils on the same items.

Items in SUBTEST TWO indicating possible differential validity as indicated by Rudner & Convey p-values analysis: TID-45^0.

ITEM NUMBER	P-VALUE DIFFERENCE
10 *	0.31341
18 #	0.27925
26 #	0.20874
31 ~	0.20857
33 *	0.21482
35 *	0.33678
36 *	0.31351
37 *	0.22594
41 #	0.22280

* = Selected items common to all four statistical strategies.

= Selected by a total of three strategies absent SSX^2.

~ = Selected by a total of two strategies absent ICC-1, and M-H.

TABLE 7 A

READING (SUBTEST TWO)

The French bilingual/English fluent pupils were examined for item responses compared with the monolingual English pupils on the same items.

Items in SUBTEST TWO indicating possible differential validity as indicated by the Rasch one parameter latent trait model: ICC-1. This subtest contains 50 items.

ITEM NUMBER	Z SCORE
10 *	-1.63416
18 #	-1.41807
26 #	1.45418
33 *	-1.33961
35 *	1.46538
36 *	1.42715
37 *	-1.39973
41 #	1.37752

* = selected items common to all four statistical strategies.

= selected by a total of three strategies: absent SSX^2.

TABLE 8 A
READING (SUBTEST TWO)

The French bilingual/English fluent pupils were examined for item responses compared with the monolingual English pupils on the same items.

Items in SUBTEST TWO indicating possible differential validity as indicated by the Mantel-Haenszel Procedure: M-H. This subtest contains 50 items.

ITEM NUMBER	CHISQ
10 *	0.023
18 #	0.033
26 #	0.070
33 *	0.102
34 *	0.103
35 *	0.021
36 *	0.031
37 *	0.087
41 #	0.035

* = selected items common to all four statistical strategies.

= selected by a total of three strategies absent SSX^2

Additionally eight items selected from the random digits table [74] from the remaining 42 items in this subtest were 36, 46, 35, 39, 49, 42, 01, These sixteen items were then divided into three groups and assigned randomly to the eight experts who had agreed to take part in this segment of this research. Note that expert number 3 served in two groups as nine were needed and only eight were avaialble.

Group 1: 3, 7, 10, 36, 46, 35; experts 4, 3, 7.
Group 2: 12, 19, 22, 39, 49, 42; experts 5, 2, 6.
Group 3: 23, 31, 01, 16; experts 8, 1, 3.

---

[74] Gene V. Glass, and Kenneth D. Hopkins, Statistical Methods in Education and Psychology Second Edition, (Englewood Cliffs, New Jersey: Prentice-Hall, 1984), 528.

Appendix D

Mathematics And Reading Questions At Issue

THE ACTUAL MATHEMATICS QUESTIONS SELECTED:

The following information, provided in order that one can understand the item documentation categories inherent in the MEA mathematics test, was provided by The Department of Education with the actual test items:

1. Three types of Process Categories
    a.) Procedural
    b.) Conceptual/Knowledge
    c.) Problem Solving
2. Five types of Content Categories
    a.) Numbers and Numeration
    b.) Variables and Relations
    c.) Geometry
    d.) Measurement
    e.) Problem-Solving Skills

The mathematics items at issue:

Selected item 3. Compute: 3 5/7 + 5 3/4=

E. 8 8/28

F. 8 8/11

G. 8 13/28

H. 9 13/28

The correct answer for item number 3 was distracter "H." The test writers' process objective for this question was to determine the pupil's ability to handle mathematical procedures, while their content objective was to determine

if the pupil could handle numbers and numeration concepts.

Item 3 favored the French pupils. French Mean: .50; SDev: .50. English Mean: .41; SDev: .49.

Selected item 7. Round 78,648 to the nearest hundred.

E. 79,000

F. 78,700

G. 78,650

H. 78,600

The correct answer for item 7 was distracter "H." The process objective was to determine the pupil's mathematical conceptual knowledge, while the content objective was to test the pupil's ability to handle numbers and numeration concepts.

Item 7 favored the French pupils. French Mean: .80; SDev: .40. English Mean: .72; SDev: .44.

Selected item 10. Last week, Karen worked 3 3/4 hours on Tuesday, 3 1/2 hours on Thursday, and 5 1/4 hours on Saturday. If Karen makes $4.20 per hour, how much did she earn last week?

A. $52.50

B. $48.30

C. $47.25

D. $46.20

The correct answer for item number 10 was distracter "A." The process objective was to determine the pupil's mathematical problem solving abilities, while the content

objective was to determine the pupil's ability to handle numbers and numeration concepts.

Item 10 favored the French pupils. French Mean:. 62; SDev: .48. English Mean: .53; SDev: .50.

Selected item 12. If m=3 and q=5, what does 3m + 4q equal?

A. 7

B. 15

C. 29

D. 78

The correct answer for item number 12 was distracter "C." The process objective was to determine the pupil's ability to handle mathematical procedures, while the content objective was to determine if the pupil could handle mathematical variables and relations.

Item 12 favored the English pupils. English Mean: .59; SDev: .49. French Mean: .51; SDev: .50.

Selected item 19. Compute: 2 2/3 x 3 2/5 =

E. 6 4/15

F. 6 4/8

G. 8 13/15

H. 9 1/15

The correct answer for item number 19 was distracter "H." The process objective was to test the pupil's mathematical procedures ability, while the content objective was to determine if the pupil could handle numbers and numeration concepts.

Item 19 favored the French pupils. French Mean: .33; SDev: .47. English Mean: .26; SDev: .44.

Selected item 22. Which figure [described] below has the greatest area?

A. 4 x 8 rectangle

B. 3 x 9 rectangle

C. 5 x 7 rectangle

D. The areas of these figures are the same.

The correct answer for item number 22 was distracter "C." The process objective was to determine the pupil's ability to handle mathematical procedures, while the content objective was to test the pupil's knowledge of geometry.

Item 22 favored the English pupils. English Mean: .37; SDev: .48. French Mean: .30; SDev: .46.

Selected item 23. Which of the following shows the numbers ordered from **least** to **greatest**?

E. 0.004, 0.06, 0.2, 0.5, 0.28

F. 0.2, 0.004, 0.5, 0.06, 0.28

G. 0.004, 0.06, 0.2, 0.28, 0.5

H. 0.004, 0.28, 0.06, 0.5, 0.2

The correct answer for item number 23 was distracter "G." The process objective was to determine the pupil's knowledge of mathematical concepts and knowledge, while the content objective the was to determine if the pupil could handle numbers and numerations concepts.

Item 23 favored the French pupils. French Mean: .46;

SDev: .50. English Mean: .39; SDev: .49.

Selected item 31. George lives 215 miles east of Powell. Allison lives 123 miles west of Powell. How far apart do George and Allison Live?

E. 92 miles

F. 112 miles

G. 192 miles

H. 338 miles

The correct answer for item number 31 was distracter "H." The process objective was to test the pupil's ability to solve problems, while the concept objective was to determine if the pupil could handle numbers and numeration concepts.

Item 31 favored the English pupils. English Mean: .54; SDev: .50. French Mean: .44; SDev: .50.

THE ACTUAL READING QUESTIONS SELECTED;

The following information, provided in order that one can understand the item documentation categories inherent in the MEA reading test, was provided by The Department of Education with the actual test items:

1. Three types of Process Categories
    a.) Comprehension
    b.) Management/Strategies
    c.) Management/Reference Skills
2. Three types of Passage Types
    a.) Literary
    b.) Content
    c.) Practical
3. Passage Length
    a.) Long Passages
    b.) Short Passages

The reading items at issue: [75]

Selected item 10: The speaker in the poem "Sidewalk Racer" describes the sidewalk as "an asphalt sea" because

E. the sidewalk is wet.

F. he feels like a sailor.

G. the sidewalk is wide.

H. he feels like an automobile.

The correct answer for item number 10 was distracter

[75] The reading passages required for fully comprehending these questions are to be found in Appendix H.

"F." The process objective was to test the pupil's reading comprehension in long, literary passage.

Item 10 favored English pupils. English Mean: .63; SDev: .48. French Mean: .54; SDev: .50.

Selected item 18: Oogruk tells Russel not to go home after he leaves him on the ice because he

E. doesn't want Russel to notify the doctor who might try to interfere.

F. is afraid the villagers will blame Russel for his death.

G. wants Russel to leave on a journey of self-discovery.

H. wants Russel to kill caribou before he returns home.

The correct answer for item number 18 was distracter "G." The process objective was to test the pupil's reading comprehension in a long, literary, passage.

Item 18 favored English pupils. English Mean: .75; SDev: .44. French Mean: .66; SDev: 47.

Selected item 26: The details of the stormy setting for this passage are typical of

E. fairy tales.

F. tall tales.

G. horror stories.

H. science fiction.

The correct answer for item number 26 was distracter

"G." The process objective was to test the pupil's ability to manage reference skills in short, literary passage.

Item number 26 favored French pupils. French Mean: .79; SDev: .41. English Mean: .73; SDev: .44.

Selected item 31: The term "classic-styled" suggests Sun Shield sunglasses are designed in

A. a classical Greek or Roman style.

B. the very latest style.

C. a traditional, tasteful style.

D. a highly scientific style.

The correct answer for item number 31 was distracter "C." The process objective was to test the pupil's comprehension of a long, practical [advertisement] passage.

Item 31 favored English pupils. English Mean: .53; SDev: .50. French Mean: .47; SDev: .50.

Selected item 33: Both paragraphs 2 and 3 under the section titled "Our Incredible Offer..."

A. emphasize the money back guarantee.

B. promise immediate delivery.

C. suggest the sunglasses are in short supply.

D. suggest saving money by ordering in quantity.

The correct answer for item number 33 was distracter "C." The process objective was to test the pupils's comprehension of a long, practical passage.

Item 33 favored English pupils. English Mean: .46; Sdev: .50. French Mean: .40; Sdev: .49.

Selected item 35: The total cost for each pair of Sun Shield sunglasses is

A. $5.00

B. $5.50.

C. $6.50.

D. $7.50.

The correct answer for item number 35 was distracter "B." The process objective was to test the pupil's ability to manage reference skills in a long, practical passage.

Item 35 favored French pupils. French Mean: .55; SDev: .50. English Mean: .46; SDev: .50.

Selected item 36: In the phrase "high resolution lenses," the word resolution means

E. pledge.

F. solution

G. clarity.

H. decision.

The correct answer for item number 36 was distractor "G." The process objective was to test the pupil's comprehension in a long, practical passage.

Item 36 favored French pupils. French Mean: .57; SDev: .50. English Mean: .49; SDev: .50.

Selected item 37: The graph is probably included to try to appeal to the reader's

A. respect for data.

B. need for leisure.

C. concern for comfort.

D. desire for style.

The correct answer for item number 37 was distractor "A." The process objective was to test the pupil's management strategies in a long, practical passage.

Item 37 favored English pupils. English Mean: .47; SDev: .50. French Mean: .40; SDev: .49.

Selected item 41: These pages **most likely** came from

A. a biography

B. a history book

C. a science book

D. an almanac

The correct answer for item number 41 was distractor "B." The process objective was to test the pupil's ability to manage reference skills in a short, practical passage.

Item 41 favored French pupils. French Mean: .75; SDev: .44. English Mean: .68; SDev: .47.

Appendix E

H. Johnson Nenty, Ph.D., Methodology

Massad (1978) listed three requirements that a test should meet before it could be regarded as a valid instrument for cross-cultural comparison. These requirements are "(1) functional equivalence (the test has to measure the same attribute on a similar scale); (2) score equivalence (the test has to measure this attribute on a similar scale); (3) item equivalence (the same requirement as score equivalence, but applied on the item level"[76]

Following from these requirements, Nenty determined that the validation of a test for cross-cultural research should proceed by checking:

> (1) how well the test measures, on the same scale, the same attribute for the different cultural groups concerned; and
>
> (2) how similar is the behaviour (sic) of each of the items in measuring on the same scale, the attribute for each of the groups.[77]

Nenty did an exhaustive study of item-bias detection methods discovering that many methods are available for use in testing for one or both of the above requirements for any test intended for cross-cultural application. Such methods had been reviewed by Jensen (1980), Merz and Rudner (1978), and compared by Ironson & Subkoviak (1979), and Rudner et.

---

[76] Nenty, 3.

[77] Ibid.

al. (1980). [78]

Nenty confirmed that:

> Studies done to establish the validity of these [above listed] methods have demonstrated that they do not select items at random (Ironson & Subkoviak, 1979; Rudner and Convey, 1978). In an artificially created situation, the amount of bias detected by SSX^2, TID, and ICC-1 methods correlated .73,.68 and .60 respectively with the actual amount of bias generated (Rudner et.al., 1980). Rudner and Convey (1978), found three methods: item characteristic curve with three parameters (ICC-3); SSX^2, and TID methods out of seven to be most promising, and observed the intercorrelation of their detected amount of bias to be .67 between SSX^2 and ICC-3, .59 between SSX^2 and TID, and .31 between ICC-3 and TID. [79]

_____C.E. Massad, International Scene, Measurement News, National Council of Measurement in Education, 9, no. 3 (1978): 3.

_____Arthur R. Jensen, Bias in Mental Testing, (New York: The Free Press, 1980).

_____W.R. Merz, and L.M. Rudner, "Bias in testing: A Presentation of selected methods." Paper presented at the annual meeting of the American Educational Research Association, Toronto (1978).

_____Gail H. Ironson, and Michael J. A. Subkoviak, "A comparison of several methods of assessing item bias, "Journal of Educational Measurement, 16 (1979): 209-225.

_____Lawrence M. Rudner, Pamela R. Getson, David L. Night, "A Monte Carlo Comparison of seven biased item detection techniques," Journal of Educational Measurement, 17 (1980): 1-10.

_____Janice Scheuneman, "A method of assessing bias in test items," Journal of Educational Measurement, 16 (1979): 143-153.

_____Rudner & Convey 1978. [source missing].

---

[78] Ibid.

[79] Ibid., 4-7.

_____Rudner et. al., 1980.

_____S.E. Whitely, and R.V. Davis, "The nature of objectivity with the Rasch Model," Journal of Educational Measurement, 11 (1974): 163-178.

_____H.J. Nenty, and T.E. Dinero, "A Cross-Cultural Analysis of the Fairness of Cattell Culture Fair Intelligence Test Using the Rasch Model," Applied Psychological Measurement 5 (1981): 355-368.

_____L.A. Marascuilo, and P.E. Slaughter, "Statistical Procedures for identifying possible sources of item bias based on X2 statistics," Journal of Educational Measurement 18 (1981): 229-249.

_____Ironson & Subkoviak, 1979.

_____Rudner and Convey, 1978, [source missing].

_____Rudner et. al., 1980.

_____Rudner and Convey, 1978, [source missing].

Appendix F

Passages For Reading Items

PASSAGES FOR READING ITEMS SELECTED BY THE STATISTICAL STRATEGIES AS INDICATING BIAS:

PASSAGE FOR READING ITEM 10:

The Sidewalk Racer
or
On the Skateboard

Skimming
an asphalt sea
I swerve, I curve, I
sway; I speed to whirring
sound an inch above the
ground; I'm the sailor
and the sail, I'm the
driver and the wheel
I'm the one and only
single engine
human auto
mobile.

"The Sidewalk Racer" from The Sidewalk Racer and Other Poems of Sports and Motion by Lillian Morrison. Reprinted by permission of Lothrop, Lee and Shepard Books.

PASSAGE FOR READING ITEM 18:

Dogsong Chapter 5

Gary Paulsen's novel Dogsong is the story of 14-year-old Russel Susskit who lives in northern Alaska on the edge of the Arctic. Russel is interested in his people's old ways which he is learning from his grandfather named Oogruk. Read Chapter 5, and then answer questions 13-20.

Russel had moved away from life in the village but he was not rebelling. He was working toward something in his mind, not away from something he didn't like. He had moved in with Oogruk, but his father knew it and approved.

There was school, of course. He was not going to school but he was learning and everybody knew that; it would have been hard to stop him trying to learn what he wanted and needed to know and so nobody tried. It would not have been polite to try it and many considered Russel old enough to know what he was doing.

Life in the village went on as it had before. Men took snowmachines out on the ice to find seals, when they could get through the leads. Other hunters took other snowmachines back into the hills and found caribou, sometimes killing six or seven to bring back for other people who could not hunt.

In the long darkness house life took on a meaning that couldn't exist in the summer. Families sometimes moved in with each other for a time, played games, fought the boredom that could come with the semi-arctic night. The village had a game room with television and it was usually crowded with both adults and children, watching the outside world.

All but Russel.

And Oogruk.

Russel hunted caribou twice more but didn't get any meat either time. He saw them at a distance, but couldn't get the sled close enough to make a stalk and a kill. On the second attempt he set the hook, left the dogs, and with the bow worked up some small creek beds but the deer saw him before he could get close enough for a shot. He took rabbits and ptarmigan home each time, using a small net Oogruk had fashioned and showed him how to use. With the net, laying it on the ground and using a long line, he lured the birds with a handful of berries. When they were on the net he flicked it closed with a jerk of his wrist and caught five and six birds at a time.

So he made meat. Light meat. That's what Oogruk called it. And it was good

meat, as far as it went. The small birds tasted sweet and were tender and soft, which suited Oogruk's poor teeth.

But the dogs needed heavy meat, heavy red meat and fat or they could not work, could not run long and hard.

And heavy meat meant deer. Caribou.

Or seal.

So it came on a cold clear morning that Russel decided to go out for seal again. It was still dark when he awakened and sat up on the floor but before he could get his pants on Oogruk was sitting up and had lighted the lamp.

"It is time for me to go out for seals again. For food for the dogs. I will go out on the ice."

Oogruk nodded. "Yes. Yes. I know that. But this time I will go with you."

Russel stopped, his bearskin pants halfway up. He looked at the old man. "To hunt seals?"

"That. And other things. There are certain things that must be done at this time and it is for an old man to do them when the time is right."

Russel waited but Oogruk said notning further. Instead he stood, slightly stiff, and feeling with his hands found clothes on the side wall. He dressed in pants and mukluks and another squirrelskin underparker. Then he took down and older outerparker, of deerskin, one with holes and worn places, and shrugged it on over his head.

"I have a good parka," Russel said. "Let me give it to you."

Oogruk shook his head. "Not this time. You keep it. You will need it and I won't. Go now and harness the dogs."

Russel finished dressing and went out for the team. They knew him now, knew him well, and greeted him with tails and barks when they saw him take the harness off the pegs. He laid the gangline out onto the snow and harnessed the team quickly, wondering why the old man wanted to go.

When the dogs were harnessed he took the weapons--two harpoons and one killing lance with a plain sharpened point--and tied them into the sled. When he turned back to the house, Oogruk had come out of the door and was looking across the ice.

His milk-white eyes stared across the ice. But he was seeing nothing. Or, Russel thought, maybe he was seeing everything.

"I smell the sea out there," Oogruk said. "It is not too far today. The ice lets the smell come across."

"The dogs are harnessed."

"I know."

"Would you drive them?"

"No. I will ride. Put me in the sled and you drive."

Russel took his hand and put him in the sled, settling him back against the crosspieces at the back. When Oogruk was settled Russel pulled the

hook and called the dogs up.

They tore away from the buildings and out across the ice. When he was away on the ice and the fire was burned out of them a bit he dragged the brake down and slowed them and looked back at the village.

Small gray buildings and caches on the dirty snow of the beach, with people here and there. Someone he did not recognize waved at him and he waved back. Dirty smoke came from chimneys and slid off with the wind and he watched as they moved away, picked up speed on the clean icesnow, until he rounded the point heading north and the buildings were gone.

He waited for some kind of sadness to come but it did not, did not, and he turned back to the sled and the dogs lined out in front and he moved them over to the right a little, using a soft "Gee," to let them know it was a gentle turn. The sea was a blue line on the horizon when they crossed the high points and could see ahead.

Oogruk said nothing, but when they got within a couple of miles of the sea and the spray smell was heavy in the cold air he held up his mittened hand to signal a halt.

"There will be seals. Watch for seals." His voice was excited, hushed but alive. "They will be on the edge of the ice. Watch for them.

Russel looked out on the edge of the ice but saw no seals. The light was half gone now and he knew that he would have to leave the sled to hunt.

"I will leave you with the dogs and go out on foot."

But now, Oogruk shook his head. "No. No. It is time to talk one more time and I must leave you. But I wanted to come out here for it because I missed the smell of the sea. I wanted to smell the sea one more time."

Russel looked down in the sled at the old man. "You're leaving me?"

"Yes. But first I must tell you what to do..."

"Where are you going?"

"It is time to leave," Oogruk said simply. "It is my time. But there is a thing you must do now to become a man. You must not go home."

"Not go home? I do not understand."

"You must leave with the dogs. Run long and find yourself. When you leave me you must head north and take meat and see the country. When you do that you will become a man. Run as long as you can. That's what used to be. Once I ran for a year to find good birds' eggs. Run with the dogs and become what the dogs will help you become. Do you understand?"v

Russel remembered now when Oogruk had said he would take a long journey. He spoke quietly. "I think so. But you, what are you to do?'

"You will leave me here

on the ice, out here by the edge of the sea."

"With respect, Grandfather, I can't do that. There is a doctor. Things can be done if something is bothering you."

Oogruk shook his head. "An old man knows when death is coming and he should be left to his own on it. You will leave me here on the ice."

"But..."

"You will leave me here on the ice."

Russel said nothing. He didn't help Oogruk, but the old man got out of the sled himself. When he was standing on the ice he motioned Russel away. "Go now."

Russel couldn't. He held back, held the sled. "I will stay with you."

"You will go." The milk-eyes looked through him to the sea, to the snow, to the line of blue that was the sky.

"You will go now."

And there was such strength in his voice that Russel knew he must go. He took the handlebar in one hand and pulled the hook, and the dogs surged away and Russel let them run without looking back. He went mile after mile, and finally he could stand it no more and he called the team around and headed back, his eyes scanning the ice in sweeps as they ran.

When they were still half a mile from where Oogruk had gotten off, Russel could see his small figure sitting on the ice and he smiled.

He would talk the old man into riding back to the village, that's all there was to it. The old man would come back and tell him more about living the old way, would sit at night and tell the stories that made the winter nights short.

But when he drew close he saw that Oogruk was sitting still. Very still. His hands were folded in his lap and his legs were stretched out in front of him and the eyes were open and not blinking with life.

Russel stopped the team before the dogs were close to Oogruk and walked ahead on foot.

Oogruk did not turn his head but stared out to sea, out past the edge of the ice where his spirit had flown, out and out. His face was already freezing and there was some blown snow in the corner of his eyes that didn't melt. Russel brushed the snow away with his mitten, a small gesture he made unknowingly, and a place in him wanted to smile and another wanted to cry. "You left too soon, Grandfather. I was coming back for you."

He stood for a time looking down at the dead old man. Then he thought of something and he went back to the sled and took the small harpoon with the ivory toggle point from the weapons lashing. He put the harpoon across Oogruk's lap so that it balanced on his knees.

"You will want to hunt seals. Use it well and make much sweet meat."

Then he went to the sled. The dogs were nervous. They smelled the death and didn't like it. The leader whined and fidgeted and was glad when Russel called them around and headed north.

Before he let them run he turned back to Oogruk one more time. "I will remember you," he said, then let the dogs go.

He would run north for a time, then cut across the ice and head northeast into the land. He had weapons and dogs and a good sled. The rest would come from the land.

Everything would come from the land.

From Dogsong by Gary Paulsen, 1985.

PASSAGE FOR ITEM 26:

FRANKENSTEIN'S AUNT

The book Frankenstein's Aunt begins a few years after Dr. Henry Frankenstein has fled from the castle where he created the monster. In this passage Henry's aunt, Hanna Frankenstein, arrives to try to restore some dignity to the family name and castle. She is met at the train station by Igor, Dr. Frankenstein's former assistant. Read to discover Hanna's reaction to the surprises that await her at the Frankenstein castle.

THEY DROVE through the narrow ravine where tree roots reared out of the mud like gnarled hands. The screeching of the cart's crooked wheels, the clip-glop of hooves in the mud, the splashing of the pouring rain and the almost incessant growl of thunder made all normal conversation quite impossible. They had to shout at each other, and Aunt Frankenstein had to shout the loudest because Igor was slightly deaf.

"Is it possible to live up here at all?" she bawled.

"Igor lives there," said Igor.

That was a surprise to Hanna Frankenstein, not altogether a pleasant surprise. She made a swift calculation in her head as to how much it would cost to send Igor to the dentist.

"Oh, so you still live there, do you, Igor? In what part of the castle. may I ask?"

"In the kitchen," screeched Igor. "It's warmest there."

The next moment, the rain-whipped air was sliced in two by a terrible flash of lightning accompanied almost simultaneously by a tremendous clap of thunder. The horse reared up on its hind legs, neighing wildly, and Igor was only just able to stop it from bolting. Aunt Frankenstein drew angrily on her cigar, which refused to burn properly in the rain; fierce raindrops came right through her umbrella and fell like mist all over her. She was soon soaked through to the skin and longed to get indoors, but when she saw the ruined castle up there in the next flash of lightning, she wondered what "indoors" would really mean.

Igor sensed her anxiety and croaked out a sound she supposed was meant to be a laugh. Hanna Frankenstein began to shiver all over. She could get furious with people who didn't give her a straight answer to her questions.

"Is it possible to live up there, or isn't it?" she snapped.

"Oh, well, you know," said Igor. "If you mend the windows and get a decent fire going, it's

probably all right I should think."

He told Aunt Frankenstein that there were bats in the chandelier in what remained of the dining room. So Hanna Frankenstein decided that it was no use being surprised by anything in the future and, in an icy calm voice, said, "And I suppose there are werewolves in the drawing room?"

"No," replied Igor, who didn't understand plurals. "I ain't seen him for ages. Maybe he's gone abroad, him too, like young Mr. Henry, you know. And I ain't seen the Count for a while, neither."

"The Count? What Count?"

"Count Dracula, of course," explained Igor, and Hanna Frankenstein's broad shoulders shuddered. She didn't like vampires, in fact disliked them as much as mosquitoes. The very thought of vampires possibly flapping around her ears while she was eating her evening sandwich depressed her.

"Are you sure those bats in the dining room aren't Count Dracula and his family?" she said. "Perhaps he's been breeding."

"Oh," said Igor. "I think I'd recognize the Count. No, them's just ordinary small bats, so there's no need to worry, Missis."

"I'm **not** worrying," said Aunt Frankenstein icily. "I'm going to put things straight in that castle again, even if werewolves and vampires are lining up in the hall."

Excerpt from Frankenstein's Aunt by Allan Rune Petterson, 1961, is reprinted by permission of Little, Brown and Company.

PASSAGE FOR ITEMS 31, 33, 35, 36,and 37:

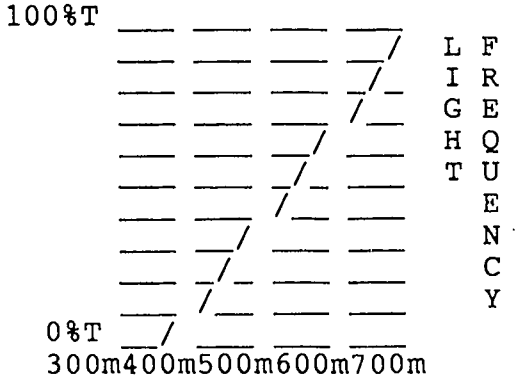
CLASSIC-STYLED

SUN SHIELD

AMBER TINTED
SUNGLASSES
With CR39
High Resolution Lenses!
$5.00

YOUR EYES ARE PRECIOUS!
DO NOT CONFUSE WITH
INFERIOR IMITATIONS
THESE ARE GENUINE CR39
HIGH RESOLUTION LENSES!

The same features found in sunglasses costing much, much more:
. Filters ultraviolet and blue light
. Highly rated CR39 gradient lenses coated to prevent scratches
. Comfortable molded nose pieces
. Polycarbonate frames with metal hinges for strength and durability
. Classic unisex styling--one size fits all



(Amber tinted lenses are highly rated to reduce UV rays and blue light.)

American Consumer has been able to obtain 100,000 Sun Shield Amber Tinted Sunglasses at a fabulous price. This enables us to offer to the public these remarkable glasses for only $5 each!

You'll Notice The Difference--
Instantly!

Our Sun Shield sunglasses have the look, the feel and features found in sunglasses selling for up to $40!

The CR39 high resolution lenses have a hard, scratch resistant coating designed to reduce glare due to reflection of surfaces such as snow, sand, or water. These lenses are perfect for fishing, sailing, water sports, tennis, golf or skiing. They're ideal for professionals who work outdoors, or under bright lights!

These are the most exciting sunglasses we've seen. Not only do they look terrific, they will protect your eyes.

Our Incredible Offer...

The moment you slip on your Sun Shield sunglasses you won't believe you purchased them for only $5...neither will anyone else who receives a pair as a gift!

Only 100,000 Sun Shield sunglasses will be offered to the public. There is a limit of 3 sunglasses per address, but if your order is mailed before December 1, 1988, you may request up to 5.

The Sun Shield sunglasses come with a 100% money back guarantee. We ship on a first-come, first-served basis with all orders shipped within 60 days. We expect a sellout--don't be disappointed. **ORDER TODAY!**

TO ORDER the Sun Shield Amber Tinted Sunglasses, send your name, address, zip code and check or money order for $5.00 plus $2.50 postage and handling for each pair ordered to: American Consumer, Box 4165, Dept KL80, Huntington Station, NJ 11746. NY, NJ, IA, IL, MI, residents add appropriate sales tax.

PASSAGE FOR READING ITEM 41.

In this exercise, you will be asked some questions about the index shown below. Do not read the entire index. Instead, read the first question and then find the answer by looking through the index. Continue until you have answered all of the questions on the page.

From The Human Achievement, Petrovich and Curtin, 1970, Silver Burdett Company

Sources Consulted

Armstrong, Liz Schevtchuk. "Education Week." Washington, D.C.: Editorial Projects in Education. (Volume X, Number 20, March 20, 1991).

"Assessing the Language Difficulties of Hispanic Bilingual Students." Reston, Virginia: ERIC Clearinghouse on Handicapped and Gifted Children. Abstract 23, August 1989. ERIC ED 321 427.

Berk, Ronald A., ed.. Handbook Of Methods For Detecting Test Bias. Baltimore: The Johns Hopkins University Press, 1982.

Bérubé, Barney. "Data Collection Report On Language Minority Children." Augusta, Maine: Department of Educational and Cultural Services Federal Projects for Language Minorities, Division of Curriculum, October 1989. A Report.

Bérubé, Barney. "Data Collection Report On Language Minority Children." Augusta, Maine: Department of Education Federal Projects for Language Minorities, Division of Curriculum, October 1991. A Report.

Cabello, Beverly. "Potential Sources of Bias in Dual Language Achievement Tests." Los Angeles, California: Center for the Study of Evaluation, California University, 1981. A Report.

Cummins, Jim. Harvard Educational Review. "Empowering Minority Students: A Framework for Intervention." (Volume 56. Number 1, February 1986).

Cummins, James. Schooling and Language Minority Students: A Theoretical Framework. Los Angeles: Office of Bilingual/Bicultural Education, California State University, Evaluation, Dissemination, and Assessment Center, 1981.

Delpit, Lisa D. Harvard Educational Review. "The Silenced Dialogue: Power and Pedagogy in Educating Other People's Children." (Volume 58. Number 3, August 1988).

Dolson, David P. The Journal of the National Association for Bilingual Education. "Bilingualism and Scholastic Performance: The Literature Revisited." (Volume 10, Number 1, Fall 1985).

Dornbusch, Sanford M.. "Bilingual Children Associated with Better Grades." East/West 20:12 (March 20, 1986).

Eisner, Elliot W.. The Educational Imagination. New York: Macmillan Publishing Company, 1985.

Ferdman, Bernardo M. Harvard Educational Review. "Literacy and Cultural Identity." (Volume 60, Number 2, May 1990).

Giles, Karen Noland. Indian High School Dropout: A Perspective. Milwaukee, Wisconsin: University of Wisconsin-Milwaukee, Midwest National Origin Desegregation Assistance Center, 1985.

Gillung, Tom B. ed. Providing Special Education Services to Limited-English-Proficient Handicapped Students. Connecticut: Department of Education, 1990.

Hakuta, Kenji. "Second Language Acquisition, Bilingual Education, and Prospects For A Language-Rich Nation." California: Stanford University, [draft], 1990.

Hambleton, Ronald K., and H. Jane Rogers. "Design of an Item Bias Review Form: Issues and Questions." Albany: New York State Education Department, 1988.

Henry, Steven L., and Floy C. Pepper. The Journal of Educational Issues of Language Minority Students. "Cognitive, Social, and Cultural Effects on Indian Learning Style: Classroom Implications." Boise, Idaho: Boise State University. (Volume 7, Special Issue, Summer 1990).

Hormel, Peter, ed.. Childhood Bilingualism: Aspects of Linguistic, Cognitive, and Social Development. New Jersey: Lawrence Erlbaum, Assoc., Inc., 1987.

Hunter, William A., ed.. Multicultural Education Through Competency Based Teacher Education. Washington, D.C.: American Association for Colleges of Teacher Education, 1977.

Ironson, Gail H., and Michael J. A. Subkoviak. "A comparison of several methods of assessing item bias." Journal of Educational Measurement. (Volume 16, 1979).

Jensen, Arthur R.. Bias in Mental Testing. New York: The Free Press, 1980.

Kavanagh, M.J., A.C. MacKinney, & L. Wolins. "Issues in managerial performance: Multitrait-multimethod analysis of ratings." Psychological Bulletin. (Volume 75, 1971), 34-49.

Kessler, C and M. Quinn. "Bilingualism and Science Problem Solving Ability." Los Angeles: Dissemination and Assessment Center, California State University. Bilingual Education Paper Series, (Volume 4, Number 1, 1980.

Kleinfeld, Judith, and Others. "The Alaska Statewide Student Testing Program: Are The Tests Biased? AK-ERA White Paper #1." Anchorage, Alaska: Alaska Educational Research Association, April 1991. A Report. ERIC ED 335 183.

Lam, Tony C.M. "Testing, Opportunity Allocation, and Asian and Pacific Americans." Honolulu: The Proceedings of a Hearing Co-Sponsored by the National Commission on Testing and Public Policy and the National Association for Asian and Pacific American Education, 1988. A Report. ERIC ED 297 058.

Linn, Robert L., ed.. Educational Measurement Third Edition. New York: Collier Macmillan, 1989.

Lucas, Tamara, Rosemary Henze, and Ruben Donato. "Promoting the Success of Latino Language-Minority Students: An Exploratory Study of Six High Schools." Harvard Educational Review. (Volume 60. Number 3, August 1990).

Lupi, Marsha H., and Joseph Yam Ting Woo. Diagnostique. "Issues in the Assessment of East Asian Handicapped Students." (Volume 14. Number 3, Spring 1989).

Maine Educational Assessment, Guide To. Augusta: Department of Education, 1988-89.

Mantel, N. and Haenszel, W. "Statistical aspects of the analysis of ^R data from retrospective studies of disease." Journal of the National Cancer Institute. (18, 1959), 719-748.

Marascuilo, L.A., and P.E. Slaughter. "Statistical Procedures for identifying possible sources of item bias based on chi-square statistics." Journal of Educational Measurement. (Volume 18, 1981), 229-248.

Massad, C.E.. International Scene. Measurement News. National Council of Measurement in Education. (Volume 9. Number 3 1978).

McBay, Shirley M., ed.. Education That Works: Quality Education For Minorities Project. Cambridge, MA: Massachusetts Institute Of Technology, 1990.

Merz, W.R. and L.M. Rudner. "Bias in testing: A Presentation of selected methods." Toronto: Paper presented at the annual meeting of the American Educational Research Association, 1978.

Miller-Jones, Dalton. American Psychologist. "Culture and Testing." The American Psychological Association, Inc.. (Volume 44. Number 2, February 1989).

Mishra, Shitala P. "Ethnic Group Bias In WISC-R Verbal Items." Anaheim, California: Paper presented at the Annual Meeting of the American Psychological Association, August 26-30, 1983. ERIC ED 236 172.

Nenty, H. Johnson. "Cross-Cultural Bias Analysis of Cattell Culture-Fair Intelligence Test." San Francisco: American Educational Research Association, April 16-20,1986. ERIC ED 274 668.

Nenty, H.J., and T.E. Dinero. Applied Psychological Measurement. "A Cross-Cultural Analysis of the Fairness of Cattell Culture Fair Intelligence Test Using the Rasch Model." (Volume 5, 1981).

Olson, David R. Harvard Educational Review. "From Utterance to Text: The Bias of Language in Speech and Writing." (Volume 47. Number 3, August 1977).

Osterlind, Steven J. "Test Item Bias." Beverly Hills and London: Sage University. Paper series on Quantitative Applications in the Social Sciences, series Number 07-030, 1983.

Rasch, G. Probabilistic Models for someApril 6, 1993 Intelligence and Attainment Tests. Copenhagen: Danish Institute for Educational Research, 1960.

Reichman, Susan, and Gloria Zyskowski. "Testing Approaches and Uses with Bilingual, Special Needs Students." Reston, Virginia: April 1988. ERIC ED 300 944.

Rhodes, Robert W. "Standardized testing of Minority Students: Navajo and Hopi Examples." St. Louis, MO: Paper presented at the 78th Annual Meeting of the National Council of Teachers of English, 1988. ERIC ED 299 587.

Rudner, Lawrence M., Pamela R. Getson, and David L. Night. Journal of Educational Measurement. "A Monte Carlo Comparison of seven biased item detection techniques." (Volume 17, 1980).

Scheuneman, Janice. Journal of Educational Measurement. "A method of assessing bias in test items." (Volume 16, 1979), 143-152.

Wechsler, D. Wechsler Intelligence Scale for Children-Revised. New York: The Psychological Corporation, 1974.

Wilkinson, Cheryl Yelich, and Wayne H. Holtzman, Jr.. "Relationships among Language Proficiency, Language of Test Administration and Special Education Eligibility for Bilingual Hispanic Students with Suspected Learning Disabilities." Texas: Department of Special Education, The University of Texas at Austin, April 1988. ERIC ED 301 604.

White House Conference on Indian Education. Washington, D.C.: 22-24 January 1992. Report in Press.

Whitely, S.E., and R.V. Davis. Journal of Educational Measurement. "The nature of objectivity with the Rasch Model." (Volume 11, 1974).

Wolfram, Walt. "Dialect Differences and Testing." Washington, D.C.: ERIC Clearinghouse on Languages and Linguistics. October 1990. ERIC ED 323 813.